PHYSICAL REVIEW

Volume 99 Number 6

PHYSICAL REVIEW

A Journal of experimental and theoretical physics

Published by the American Institute of Physics

AMERICAN PHYSICAL SOCIETY

President

S. T. Bowen
University of California
Berkeley, California

Vice-President

R. P. Winston
Princeton University
Princeton, New Jersey

Secretary

E. K. Dunnow
Bell Telephone Laboratories
Office
Columbia University
New York 27, New York

Treasurer

G. B. Pegram
Columbia University
New York 27, New York

Local Secretary
for the Pacific Coast

William A. Nierenberg
University of California
Berkeley, California

Managing Editor

S. A. Goudsmit
Brookhaven National Laboratory
Upton, Long Island
New York

Editor

S. A. Goudsmit

Associate Editors

W. A. Fowler
Kenneth Greisen
Charles Kittel
R. E. Marshak
W. K. H. Panofsky
V. F. Weisskopf

Associate Editors

G. J. Dienes
R. G. Fluty
K. G. McC
J. R. Pell
Lewi Tons
Gregor Wa

Manuscripts for publication should be submitted to S. A. Goudsmit, Brookhaven National Laboratory, Upton, Long Island, New York. Authors' institutions are requested to defray publication charges of $[illegible] per page which, if honored, entitles them to [illegible] reprints. Instructions will be sent with galley proofs.

Proof and all correspondence concerning papers in the process of publication should be addressed to the Publication Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Rates

United States and Canada [illegible]
Elsewhere [illegible]

Publication Schedule

Yearly [illegible] number and index [illegible]
Price to [illegible]
Three [illegible]

Single copies: [illegible]

General index, [illegible]

Subscriptions, [illegible] should be sent to the American Institute of Physics, [illegible] New York 22, New York.

Changes of address in the [illegible] American Physical Society should be addressed to the [illegible] other subscriptions to the American Institute of Physics.

[illegible] Prince and Lemon Streets, Lancaster, Penna. New York: [illegible]

The Physical Review is published semi-monthly on [illegible]

Entered as second-class matter March 6, 1931, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, [illegible]

Acceptance for mailing at special rate of postage provided for in paragraph [illegible] Act of [illegible] authorized June [illegible]

THE
PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 99, NO. 6 SEPTEMBER 15, 1955

Second Sound Attenuation in Rotating Helium II*

R. G. WHEELER,† C. H. BLAKEWOOD, AND C. T. LANE
Sloane Physics Laboratory, Yale University, New Haven, Connecticut
(Received June 10, 1955)

An experiment is described wherein liquid helium II is rotated in an annulus between a metal cylinder which rotates at constant speed and a concentric cylindrical cavity at rest. Second sound pulses, propagated parallel to the cylindrical axis and through a substantially uniform velocity field, are found to suffer an extra attenuation due to the motion of the fluid. If we define σ as the ratio of the received thermal pulse heights (fluid in motion *versus* fluid at rest), then for angular velocities of the cylinder ω and second sound path lengths l it appears that $\sigma=e^{-\beta\omega l}$. The coefficient β decreases somewhat with rising temperature in the range from about 1.2 to 2.0°K, having a value of $(8\pm0.8)\times10^{-6}$ rev^{-1} cm^{-1} sec at 1.39°K. The effect is discussed in the light of present knowledge concerning the nature of helium II.

INTRODUCTION

IN 1941, Kapitza[1] reported a notable series of experiments aimed at providing a better understanding of the then obscure properties of liquid helium II. Among these was one which operated substantially as follows. A glass capillary was provided with a concentric glass rod, the annular space between the pair being filled with helium II. Various constant angular velocities could be imparted to the central rod while the capillary tube remained at rest. In effect, Kapitza measured the thermal conductivity of the liquid, in the annular space, as a function of the angular velocity of the central cylinder. The result reported showed a decrease in the thermal conductivity of He II, under rotation, as compared with the conductivity for the system at rest.

This interesting experiment does not appear to have been repeated and has, in fact, received very scant notice in the numerous review articles which have been written since that time. We therefore undertook to attempt a variation on the above experiment using a very similar geometrical arrangement but measuring the second sound velocity rather than the effective thermal conductivity. The second sound is also a thermal property; it was, of course, undiscovered at the time of Kapitza's experiments, and we hazarded the guess that a positive effect might occur. After a little experimentation, however, it emerged that there was no measurable effect of rotation on the second sound velocity within a considerable range of angular velocities of the central cylinder.[2] We did, however, notice that an "extra" attenuation in the thermal pulses occurred as a function of the rotation. In what follows, we present some measurements on this rotational attenuation effect in the second sound.

EXPERIMENTAL METHOD

In order to make attenuation measurements which can be attributed only to the rotation of liquid helium, the technique used must be such that the "beam spreading" and the normal attenuation are eliminated. Also we desire to have the second sound propagated through a substantially uniform fluid velocity field. Finally we must use a signal frequency or pulse width which will give us a measurable attenuation, since the normal attenuation coefficient varies as the square of the signal frequency.

With these factors in mind, the rotation of the liquid was set up in the annulus between two coaxial cylinders, of which the inner was rotated while the outer one remained at rest. Second sound was propagated in this

* Assisted by a grant from the National Science Foundation. A preliminary account of this work was given at the New York Meeting of the American Physical Society in January, 1955 [Phys. Rev. 98, 1196(A) (1955)].
† DuPont Fellow in Physics. This work forms part of a Ph.D. dissertation.
[1] P. L. Kapitza, J. Phys. (U.S.S.R.) 4, 181 (1941).
[2] Our velocity measurements were such that a change in velocity of 0.1% could have been detected.

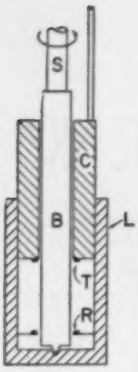


FIG. 1. Essentials of the apparatus in the cryostat.

annulus parallel to the axis of rotation. The transmitter and receiver were annular strips whose width was about one-seventh the width of the annulus between the cylinders, thus propagating in as small a region of the annulus as technically possible. These two elements remained stationary and did not rotate with the fluid. A pulse technique was dictated by two considerations. The first being the fact that it was necessary to separate out, in time, the signals which travel various discrete paths from the transmitter to the receiver. Secondly, the pulse width had to be small enough so that the attenuation could be measured for reasonable path lengths.

With this geometry and the pulse widths used, the transmitter was of the order of 30 wavelengths wide, thus appreciable beam spreading is encountered. The apparent attenuation due to beam spreading may then be eliminated by comparing signals with rotation to those with no rotation holding all other parameters constant. This of course implies that the beam spreading is not a function of the rotation of the liquid. We believe this to be so since the velocity, hence the wavelength, does not change with rotation.

TECHNICAL DETAILS

Figure 1 is a somewhat schematic and much simplified picture of our experimental arrangement. This went through several evolutions and this was the final one. The central cylinder B was brass 1.588 cm in diameter and approximately 14 cm long. The concentric cavity L was of Lucite (for optical transparency) and was 3.175 cm in internal diameter[3]; hence the annular gap was 0.793 cm wide. The second sound transmitter T and receiver R were flat annular strips of carbon resistor board approximately 1.23 mm wide. These were constructed by painting two concentric circles, of slightly different diameter, on the resistor board using silver paint. These were each connected to an electrical lead and so the current passed through the resulting annular carbon strip in a radial direction. This permitted us to use carbon strip of high specific resistivity and yet end up with an element of moderate resistance at helium temperatures. These carbon resistor elements were constructed from *IRC* resistance strip, and had a dc resistance at 1.39°K of approximately 4200 ohms. The center of each strip was located at a distance of 1.08 cm from the axis of cylinder B; i.e., they were asymmetrically placed in the annular gap formed by B and L. Each strip was provided with small coaxial cable leads[4] out through the top of the cryostat.

The brass cylinder was driven from a motor at the top of the cryostat via a stainless steel tubular shaft S the bearings (not shown) being of Kel-F plastic. The receiver was mounted in a fixed position near the bottom of the apparatus, but the transmitter being attached to the sliding Lucite cylinder C could be moved, within the cryostat, in a direction accurately parallel to the axis of B such that the distance between T and R could be set at any desired value up to 14 cm. This distance could be measured, under helium, by means of a cathetometer looking through slits in the Dewars.

The motor driving B consisted of one of a pair of Bendix synchos. This was mounted above the cryostat and coupled into the shaft in the Dewar by means of a simple magnetic clutch using Alnico magnets. Thus vacuum tightness of the cryostat was preserved without using a packing gland for shaft S. The twin syncho was driven at various required speeds by an electronically controlled, variable-speed dc motor. The average variation in speed obtained with this driving mechanism was less than 0.5 percent. Other things being equal, no difference in the results was observed for clockwise or anticlockwise rotation of B.

The transmitter was fed with single "square" electrical pulses of approximately 5- or 10-μsec duration, but due to the thermal time constants of the transmitter the second sound pulse was spread out to about 20 μsec. The electrical power in most runs was about 0.7 watt/cm^2 at the transmitter but, due to beam spreading, the effective area soon becomes that of the entire annulus rather than that of the carbon strip. The receiver was pulsed with a small dc bias current and the resistance change resulting from the received second sound heat

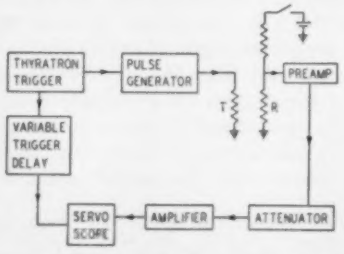


FIG. 2. Block diagram of the electronic circuitry.

[3] All measurements at room temperature.

[4] R. G. Wheeler and S. J. Plaskon, Rev. Sci. Instr. **26**, 404 (1955).

pulse displayed, in the usual way, on an oscilloscope with a fast, triggered sweep and photographed with a Praktiflex 35-mm camera. The electronics was based on that employed previously by Fairbank and coworkers[5,6] in this laboratory, but due to beam spreading we had to pay great attention to the signal-to-noise ratio.

Figure 2 shows a block diagram of our arrangement. In particular the preamplifier had to be designed and constructed with great care. We finally made use of a special Western Electric 417A triode[7] for the input tube. These triodes when placed in a cascode circuit gave an equivalent input noise resistance of about 80 ohms. With this arrangement our signal to noise ratio, under optimum conditions, amounted to better than 40:1.

OBSERVATIONAL PROCEDURE

With the distance between T and R set at some chosen value and the temperature set at some given point (as judged on an Octoil-S manometer) we photographed a series of pulses, usually about 8 frames, with B at rest. Cylinder B was then set in constant rotation, at some given value, determined with a tachometer on the driving sychro. This invariably caused a rise in the temperature of the helium bath for obvious reasons. The bath temperature was then brought back to its previous value by manual manipulation using the oil manometer as the criterion.[8] When conditions were steady a second series of about 20 pictures was taken. The motor driving B was then stopped, the bath temperature again brought back to its proper value, and a third set of approximately 8 "stills" taken. Thence either a new speed, a new path length or a new temperature was chosen depending on which parameter we were varying, and the above process repeated. Thus each point on our graphs represent the measurement of about 36 pictures of which roughly half were for B at rest. Occasionally, however, we took approximately three times as many pictures, per observational point, in an effort to achieve better experimental accuracy.

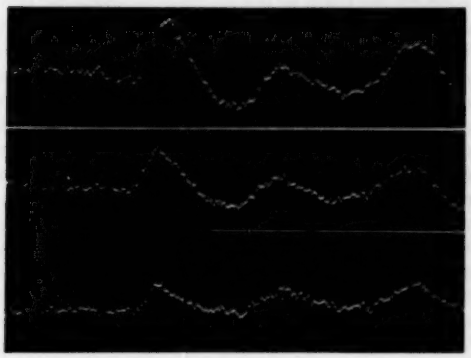


FIG. 3. Three photographs of oscilloscope traces showing the attenuation of second sound by rotation. Time (250 μsec sweep) is proceeding from left to right. In each case, the left hand pip is that of the direct second sound pulse followed, in order, by echo 2 and echo 3. The top trace is for the cylinder B at rest ($\omega=0$); the middle trace for $\omega=7.2$ rev/sec and the bottom trace for $\omega=13.2$ rev/sec. In all three cases the temperature is 1.39°K, the path length for the direct pulse $l=5.70$ cm and the input electrical pulse is of 10-μsec duration. Echo 1 is unresolved but some of our photographs show it as a small pip on the rear edge of the direct pulse.

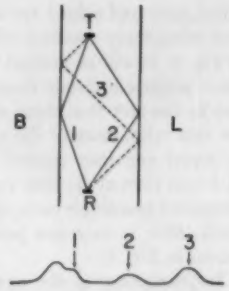


FIG. 4. Illustrating the paths of three possible second sound echoes plus a reproduction of a typical oscilloscope trace showing their resolution in time.

The reason for the great care exercised in the temperature control was, of course, due to the inherent temperature dependence of the sensitivity of the receiver. A separate resistance *versus* temperature measurement (using bath vapor pressure in the usual way) on our receiver showed, however, that a bath-temperature variation of a few millidegrees was entirely negligible in its effect on the size of our recorded pulse.

From each set of 36 pictures we then measured the ratio of the received pulse heights (cylinder B in motion as against cylinder B at rest). This quantity we call the "relative pulse height" (σ) in what follows. Figure 3 shows a sample of our pictures.

It has previously been stated that a pulse technique was dictated by our geometry. This is due to the fact that we observe many second sound echoes caused by reflections at B and L—some of our photographs showed as many as 6 echoes in addition to the direct beam. A continuous wave technique, such as that employed by Hanson and Pellam[9] to study the normal attenuation in He II, would here be useless since it would not resolve any echo. As a matter of fact our pulse technique did not resolve them all. Thus, referring to Fig. 4, the paths for the first three possible echoes are indicated. Due to the asymmetry of our transducer system in the annular gap, signals for the

[5] E. A. Lynton and H. A. Fairbank, Phys. Rev. **80**, 1043 (1950).
[6] J. C. King and H. A. Fairbank, Phys. Rev. **93**, 21 (1954).
[7] Our thanks are due Dr. W. A. Tyrrell of Bell Telephone Laboratories for his gift of several of these triodes.
[8] In order to check the validity of this procedure we made a dummy run using the receiver as an ordinary resistance thermometer. The result showed that we could reset the temperature with the oil manometer, when B was rotating, to within 3 millidegrees of its value when B was at rest.
[9] W. B. Hanson and J. R. Pellam, Phys. Rev. **95**, 321 (1954).

direct path and echo 1 are not completely resolved, the first completely resolved echo has a path labelled as 2 in Fig. 4. It will be noticed that the next echo (path 3) has a larger amplitude than that due to path 2. This is due to the fact that there are two equal paths possible for this echo, namely the one labeled 3 in Fig. 4 plus an equal one (not shown) created by reflections first at L and then at B; thus twice the transmitted energy, compared to a single path, will be received. On the other hand there is only one possible path for echo 2, that shown in Fig. 4.

In presenting, in the next section, our experimental results we have found it convenient to use the following nomenclature: l=least distance in cm between transmitter T and receiver R. (This we call the "path length.") S_0=height of the recorded second sound pulse (via path l) when cylinder B is at rest. S_ω=height of the same pulse when cylinder B has a constant angular velocity. ω=angular velocity of cylinder B in rev/sec. $\sigma=S_\omega/S_0$, a dimensionless number called the "relative pulse height."

RESULTS

Figure 5 is a plot of σ *versus* l for $\omega=15$ rev/sec at a temperature of 1.39°K. The relationship is seen to be accurately exponential, albeit the curve is a rather flat one. In order to determine the form of the function $\sigma=F(\omega,l)$ at constant temperature we made several runs, of the above nature, for several different values of the speed ω. A plot of logσ *versus* l (Fig. 6) yields a family of straight lines all extrapolating to the value $\sigma=1$ at $l=0$. A plot of the slope of these lines *versus* ω (Fig. 7) yields another linear relationship. From this it appears, therefore, that the functional relationship between σ, ω, and l is

$$\sigma=\epsilon^{-\beta\omega l},$$

where β is a constant which we shall hereafter call the rotational attenuation coefficient.

All the above analysis has been carried out at one temperature, namely, 1.39°K and the question as to

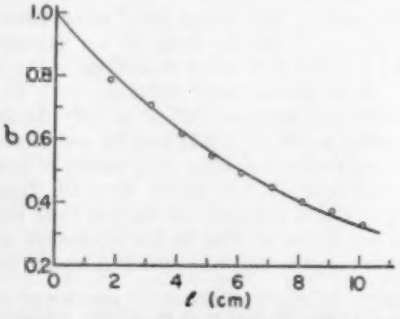


FIG. 5. Relative pulse height (σ) *versus* path length (l) at 1.39°K for $\omega=15$ rev/sec. The electrical input pulse was 10 μsec duration. The circles are the experimental points and the solid curve is a plot of $\sigma=\epsilon^{-\beta\omega l}$ with $\beta=7.6\times10^{-3}$ rev^{-1} cm^{-1} sec.

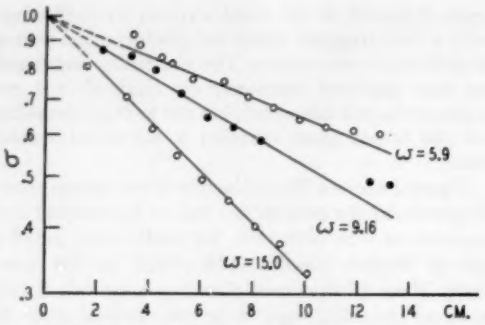


FIG. 6. Semilog plot of relative pulse height (σ) *versus* path length (l) at 1.39°K for three different speeds of cylinder B. The electrical input pulse was of either 5 or 10 μsec duration

whether β is temperature dependent remains, for the moment, open. At 1.39°K, and for the geometry used in the experiment, $\beta=(8\pm0.8)\times10^{-3}$ rev^{-1} cm^{-1} sec.

In order to test the truth of the above hypothesis somewhat further, Fig. 8 shows a plot of σ *versus* ω at $l=5.70$ cm and $T=1.39$°K. The curve is, within our accuracy, an exponential, as expected, yielding, for this particular run, a $\beta=8.3\times10^{-3}$ rev^{-1} cm^{-1} sec. To answer the question concerning the possible temperature dependence of the coefficient β we performed the following kind of experiment. With l and ω fixed, at some chosen values, we measured σ as a function of temperature in the range from about 1.2 to 2.0°K. Assuming the truth of the above relationship between σ, ω, and l, we then compute β at each temperature. The result is shown in Fig. 9.

We found it necessary to use two different transmitters to cover the above temperature range. The reason for this lies in the fact that, as the lambda point is approached, the specific heat of liquid helium increases sharply. Hence more power must be employed around 2°K than was used at, say, 1.2°K in order to produce the same temperature pulse height in the two cases. Our original transmitter, whose resistance at helium temperatures was about 4000 ohms, had too high a resistance to give sufficient pulse amplitude for accuracy beyond about 1.8°K. We therefore constructed a second one, identical in dimensions, but with a lower resistance so that more power could be fed into the helium from our square-wave generator.

This necessary increase in power, at the higher temperatures, leads to some difficulties connected with the formation of shock wave effects. As is well known,[10] shock effects are disturbing to both velocity and attenuation measurements in second sound. We were careful, therefore, to limit our powers to values such that measurable signals were produced without the formation of noticeable shock wave effects. Nevertheless, above 2°K, we were unable to produce usable

[10] K. R. Atkins, Phil. Mag., Suppl. **1**, 169 (1952).

signals at any power available to us. We attribute this largely to the rapid increase in the natural attenuation which is known to occur above about 2°K,[9] coupled with the decrease in sensitivity of our receiver with rising temperatures.

DISCUSSION

A considerable effort, both theoretical and experimental, has been devoted to what we have called the "normal" attenuation of the second sound in helium II. The theory is principally due to Khalatnikov[11] and in this analysis the attenuation is caused by three types of dissipative mechanism, e.g., (i) the viscosity of the normal component of the liquid, (ii) the so-called second or bulk viscosity and (iii) dissipation due to diffusive thermal flow in the normal component. These various factors are computed by Khalatnikov on the basis of the assumed roton and phonon spectra forming the normal component. This theory seems to be in substantial agreement with the latest experimental observations of Hanson and Pellam[9] and Atkins and Hart,[12] although some of the essential parameters in the theory require empirical determination for the numerical values.

Accordingly it would seem that the observed rotational attenuation requires some new dissipative mechanism in addition to those listed above, but we are unable to hazard a guess as to what the precise nature of this might be. In this connection, however, an interesting experiment due to Hollis-Hallett[13] gives results which probably have some bearing on ours. Hollis-Hallett has adapted the rotating cylinder viscometer to the measurement of the viscosity of the normal component of helium II. The geometry of this device is very similar to that employed in our experiments except for the difference that the outer cylinder rotates and the central one is at rest. In a classical fluid the torque on the fixed cylinder is found to be directly proportional to the velocity of the rotating one which

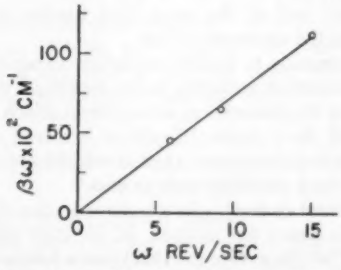


FIG. 7. A plot of the slopes of the lines in Fig. 6 *versus* the angular velocity ω.

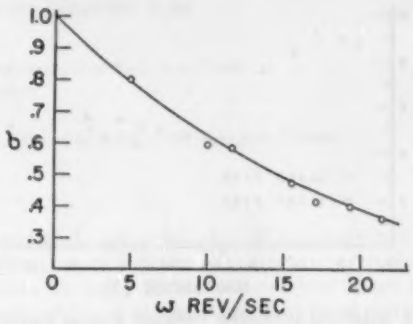


FIG. 8. Relative pulse height (σ) *versus* angular velocity of cylinder B at 1.39°K for a constant $l=5.7$ cm. The electrical input pulse was of 10-μsec duration. The circles are the experimental points and the solid curve is a plot of $\sigma=e^{-\beta\omega l}$ with $\beta=8.3\times10^{-4}$ rev^{-1} cm^{-1} sec.

is, of course in agreement with predictions of the Navier-Stokes equation which in turn shows that the proportionality constant contains the first viscosity. In helium II, however, Hollis-Hallett finds that the torque is not so linearly related even for velocities of as little as 0.1 cm/sec. The observed torque, at any velocity, was larger than that required by a linear law or, stated another way, there appears to be an "extra viscosity" over and above the classical one such as our experiment would also seem to require.

The hydrodynamics of helium II is based on the classical Navier-Stokes equation, for both normal and superfluid, with added terms. These include the observed pressure gradient in the superfluid induced by heat flow (fountain effect) and, in addition, the postulated "mutual friction" between normal and superfluid of Gorter and Mellink.[14] At first sight this latter term might seem to account for the extra viscosity but, as Hollis-Hallett shows,[13] it does not. The prediction of the adjusted Navier-Stokes equation is still for a torque linear with the velocity. It appears therefore that the hydrodynamic equation for He II are still incomplete.[15]

The apparent failure of the adjusted Navier-Stokes equation, discussed above, could also have some bearing on our experiment in another way. Thus it appears from our work that the transverse velocity of presumably the normal fluid (perhaps also the superfluid) increases the normal attenuation of second sound. But we do not, at the moment, know the magnitude of these velocities with any degree of certainty—the best we can do is to compute the velocity profile in the annulus from the Navier-Stokes equation thus treating the helium as if it were a classical liquid. This gives for the angular velocity of the fluid $\omega(r)$ at distance r from the

[11] I. M. Khalatnikov, J. Exptl. Theoret. Phys. (U.S.S.R.) **20**, 243 (1950); **23**, 34 (1952).
[12] K. R. Atkins and K. H. Hart, Phys. Rev. **92**, 204 (1953).
[13] A. C. Hollis-Hallett, Proc. Cambridge Phil. Soc. **49**, 717 (1953).
[14] C. J. Gorter and J. H. Mellink, Physica **15**, 285 (1949).
[15] In addition to the Gorter-Mellink formulation of the hydrodynamic equations there are a number of others which, however, have been subjected to less experimental testing than these above. For details see J. G. Daunt and R. S. Smith, Revs. Modern Phys. **26**, 172 (1954).

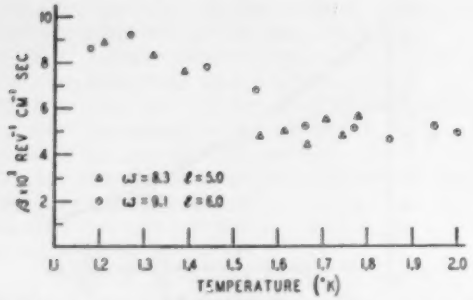


FIG. 9. Rotational attenuation coefficient β *versus* temperature.

axis of rotation of cylinder B:

$$\omega(r)=\omega\left(\frac{a^2}{b^2-a^2}\right)\left(\frac{b^2}{r^2}-1\right), \tag{1}$$

wherein a and b are the radii of cylinders B and L respectively and ω is the angular velocity of cylinder B.[16] It may well be, however, in lieu of an actual velocity measurement in the annular gap, that the above relation does not express the true facts. We are in a position to make some test of this by considering the attenuation of the echoes indicated in Fig. 4, and from our data for a 5-cm path length have measured the relative attenuation of echoes 2 and 3 as a function of the average fluid velocity,[17] computed from relation (1) above. The result, together with that for the direct path, are shown in Fig. 10, and in this graph the flat exponential has been drawn as a straight line. It will be seen that the result appears reasonable. Thus the curves for the direct pulse and echo 3, which pass through average velocity fields of the same order of magnitude, are only slightly different but the attenuation effect in echo 2 which, according to calculation, passes through a smaller velocity field is much reduced. We were unable to make measurements on echo 1 since it is not resolved in our data. This does not, of course, establish the correctness, or otherwise, of the velocity profile as deduced above. It does, however, indicate that the velocity cannot be nearly uniform across the annulus.

From this it appears probable that the coefficient β, as defined by us, would depend on the position of the transducer system in the annular gap in which the helium rotates. The figure $(8\pm0.8)\times10^{-3}$ rev^{-1} cm^{-1} sec at 1.39°K thus refers to a position such that the center of the system is distant 1.08 cm from the axis of the cylinder B. We made no measurements to test this point. In addition we have no information about the possible dependence of β on frequency. Our method is such that we cannot substantially alter the width of the input pulse (and thus alter the fundamental pulse frequency) and at the same time resolve the direct pulse and the numerous echoes.

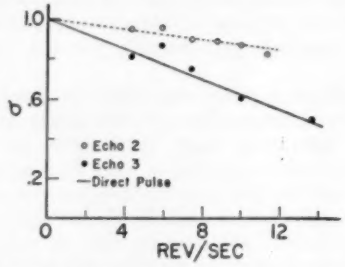


FIG. 10. The relative pulse height (σ) *versus* angular velocity of central cylinder for the direct beam and two echoes. Temperature 1.39°K with a 5-cm direct beam path length.

It is tempting to ascribe our results to some kind of turbulent motion occurring in the superfluid, especially in view of the geometrical arrangement which we have employed. In a classical liquid it is known that the motion is turbulent for all angular velocities of the inner cylinder for a geometry such as ours.[18]

Note added in proof.—In a communication appearing after this paper was written, H. E. Hall and W. F. Vinen [Phil. Mag. **46**, 546 (1955)] have briefly reported what appears to be an entirely different kind of experiment also showing an effect of rotation on the second sound attenuation.

[16] The transient term, for our geometry, which multiples the expression for the steady-state velocity distribution is given by

$$\{1-\exp(-\lambda^2\mu t/\rho)\},$$

where ρ and μ are the density and viscosity and λ is a geometrical factor obtained from the boundary conditions ($\lambda^2\cong16$ cm^{-2}). Using the normal component density and viscosity at 1.4°K, this "time constant" is of the order of 40 seconds. We looked for such a time effect; however our experimental procedure limited us to variations in time occurring about 20 seconds after the system had started rotating. No such transient effect in the amplitude of the second sound pulse with rotation was observed. It must be stated, however, that the solution of the Navier-Stokes equation which yields the above transient term is probably too approximate. We have to neglect the centripetal force and make the even worse assumption that the velocities are sufficiently small to neglect second order terms.

[17] Referring to Fig. 4 it is apparent that whereas the direct pulse passes through a substantially constant velocity field, echoes 2 and 3 (if relation 1 is true) do not. We therefore define an average velocity field $\langle\omega\rangle_{\text{av}}$ for these cases as

$$\langle\omega\rangle_{\text{av}}=\frac{1}{r_0}\int_0^{r_0}\omega(r)dr,$$

where r_0 is the path length, between transmitter and receiver, for the echo. This quantity can be computed from the known geometry of the apparatus. Although the rotational attenuation depends also on the length of path it turns out that the paths for echoes 2 and 3 are, respectively, only about 2 and 5 percent longer than the direct pulse path in our apparatus. Hence this fact is ignored. The average velocities as computed from relation (1) are as follows:

direct pulse: $\langle\omega\rangle_{\text{av}}=0.376\omega$,
echo 1: $\langle\omega\rangle_1=0.636\omega$,
echo 2: $\langle\omega\rangle_2=0.152\omega$,
echo 3: $\langle\omega\rangle_3=0.333\omega$.

[18] H. Lamb, *Hydrodynamics* (Cambridge University Press, Cambridge, 1916), fourth edition, p. 655.

Virial Series of the Ideal Bose-Einstein Gas

H. S. SHAPIRO
Institute of Mathematical Sciences, New York University, New York, New York
(Received April 22, 1955)

B. Widom has conjectured that the radius of convergence of the virial series of the ideal Bose-Einstein gas is infinite. The present note shows this to be false.

WIDOM[1] considers the virial series of the ideal Bose-Einstein gas, that is, the power series

$$x_2=\sum_{n=1}^{\infty} a_n x_1^n, \tag{1}$$

where

$$x_s=\sum_{1}^{\infty} n^{-(s+\frac{1}{2})}y^n,$$

and shows that R, the radius of convergence of (1), is ≥ 0.257. He conjectures further that R is infinite, i.e., that (1) converges for all x_1. We show here that this conjecture is false.[2]

Our first step follows Newman. Let

$$x_2=f(x_1), \quad f \text{ entire.} \tag{2}$$

Then $dx_2/dy=f'(x_1)(dx_1/dy)$ whence, since $y(dx_2/dy)=x_1$, we have

$$\frac{dy}{y}=\frac{f'(x_1)}{x_1}dx_1=\left[\frac{1}{x_1}+g(x_1)\right]dx_1,$$

where g is entire. Integrating, we get $y=x_1e^{G(x_1)}$, where G is entire. Thus (2) implies that y is an entire function of x_1, and it suffices to show (writing z for y): *The function-element defined by the power series*

$$u(z)=\sum_{n=1}^{\infty}\frac{z^n}{n^{\frac{1}{2}}}, \quad |z|<1 \tag{3}$$

is not inverse to an entire function.

To show this we must obtain the analytic continuation of $u(z)$ outside the unit circle. This is best done using the integral representation

$$u(z)=\frac{4}{3\sqrt{\pi}}z\int_0^{\infty}\frac{dt}{\exp(t^2)-z},$$

from which one easily deduces: All singularities of the complete analytic function $U(z)$ (gotten by continuing $u(z)$ onto its entire Riemann surface) lie above $z=0$ and $z=1$. $z=1$ is an algebraic branch point of order one on every sheet, and $z=0$ is a logarithmic branch point on every sheet but the original. When $u(z)$ is continued once around $z=1$ it passes into the new function element

$$u_1(z)=u(z)-4i(\pi\log z)^{\frac{1}{2}}. \tag{4}$$

From this information we can complete the proof. Let γ_1 be a path going from $z=0$ once around $z=1$ and back to a point z of $0<|z|<1$, and let γ_2 be a path going from $z=0$ once around $z=1$, then once around $z=0$, then again around $z=1$ and back to a point z of $0<|z|<1$. The result of continuing $u(z)$ along γ_1 is given by (4), from which we deduce that the result of continuing $u(z)$ along γ_2 is the element

$$u_2(z)=u(z)-4i(\pi\log z)^{\frac{1}{2}}-4i[\pi(\log z+2\pi i)]^{\frac{1}{2}}. \tag{5}$$

Suppose now that γ_1 terminates in the point z_1, and γ_2 in z_2, and $u_1(z_1)=u_2(z_2)=\alpha$. From (4) and (5) we see that $z_1\neq z_2$. Then, if Γ_1 and Γ_2 are the images of γ_1 and γ_2 in the U-plane, the result of continuing the function $F(U)$ [inverse to $U(z)$] from $U=0$ to $U=\alpha$ along the paths Γ_1 and Γ_2 is to obtain two distinct determinations at $U=\alpha$, namely z_1 and z_2. Thus the inverse function $F(U)$ could not even be single-valued, and *a fortiori* not entire.

To complete the proof, we prove the existence of γ_1 and γ_2 with the desired properties. By letting γ_1 run around the origin in all possible ways after its return to $0<|z|<1$, $w=(\log z)^{\frac{1}{2}}$ takes on all values in the sector S: $\frac{1}{4}\pi<\arg w<\frac{3}{4}\pi$ and therefore $u_1(z)$ takes on all values

$$u(\exp(w^2))-4i\pi^{\frac{1}{2}}w. \tag{6}$$

Similarly, by letting γ_2 run around the origin in all possible ways after its final return to $0<|z|<1$, $u_2(z)$ takes on all values

$$u(\exp(w^2))-4i\pi^{\frac{1}{2}}w-4i\pi^{\frac{1}{2}}(w^2+2\pi i)^{\frac{1}{2}}. \tag{7}$$

It remains only to show that the sets of values assumed by the functions (6) and (7), respectively, as w ranges over S, are not disjoint. Since $|u(\exp(w^2))|<\zeta(3/2)$, and since $(w^2+2\pi i)^{\frac{1}{2}}$ differs from w by a bounded function in S, it suffices to prove the same for the values

[1] B. Widom, Phys. Rev. **96**, 16 (1954).

[2] This was proved recently by Dr. D. J. Newman of the Republic Aviation Corporation, who showed $R\leq 64$. Newman's proof (not published) uses a generalization of Picard's theorem on entire functions, which we are able to avoid.

Since the submission of this note, the author has been informed of prior work of Dr. W. H. J. Fuchs of Cornell University, who has obtained the more precise result $12.56\cdots\leq R\leq 27.73\cdots$. The methods of Fuchs and the present author are essentially the same, our presentation (of the more limited result) being, however, more compact and transparent.

assumed by functions

$$\phi_1(w)-4i\pi^{\frac{1}{2}}w \tag{6a}$$

and

$$\phi_2(w)-8i\pi^{\frac{1}{2}}w, \tag{7a}$$

where ϕ_1 and ϕ_2 are regular and bounded in S. Because the correspondence $w\leftrightarrow 2w$ maps S onto itself, we may without loss of generality replace the 8 in (7a) by a 4, and (dividing by $4i\sqrt{\pi}$ and writing $\psi_k=\phi_k/4i\sqrt{\pi}$) we have only to prove that the value-sets of the functions

$$w-\psi_1(w), \tag{6b}$$

and

$$w-\psi_2(w) \tag{7b}$$

are not disjoint, where ψ_1 and ψ_2 are regular and bounded in S. But this is immediate: for, if $|\psi_k|\leq M$, $k=1, 2$ then by Rouche's theorem both functions (6b) and (7b) assume the value $M\sqrt{2}i$ inside the circle of center $M\sqrt{2}i$ and radius M.

Remark.—If K_i denotes the maximum of $|U(z)|$ on γ_i $(i=1, 2)$, and $\tau=\max(K_1,K_2)$, the above reasoning gives the more precise result that $F(U)$ has a branch point in the circle $|U|\leq\tau$, whence τ is an upper bound for the radius of convergence R of the virial series. It has not been deemed worth while to make a numerical estimate (which would be a straightforward task) because in any case the method is too crude to answer the interesting question of whether (1) converges at $x_1=\zeta(3/2)=2.612\cdots$, the value of the dimensionless density for which condensation is known to occur.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Turbulence Spectrum in Chandrasekhar's Theory*

JOSEPH W. CHAMBERLAIN AND PAUL H. ROBERTS
Yerkes Observatory, University of Chicago, Williams Bay, Wisconsin

(Received April 4, 1955)

The results of Chandrasekhar's recent theory of turbulence are transformed in this paper from ordinary space (used exclusively in his paper) to wave-number space. Consideration is limited to the case of stationary, homogeneous, and isotropic turbulence. A "time-dependent spectrum" is defined in terms of the scalar product of eddy velocities at two different times; this spectrum is related to Chandrasekhar's time-dependent correlation function by a Fourier transform, as in the conventional, time-independent case. For infinite Reynolds number the spectrum is obtained directly by transforming the correlation function into wave-number space; the spectrum is given by a much simpler expression than is the corresponding correlation function.

Consequently, an approximate (linearized) form of the *differential* equation for the correlation has been transformed to wave-number space, and this equation is readily solvable, even for a finite Reynolds number. For $k>1/\nu$ the spectrum vanishes, and this cutoff at large wave-numbers is interpreted as the disintegration of turbulence into laminar flow at dimensions sufficiently small for viscosity to dominate over the inertial transfer of energy.

Finally the general, nonlinear correlation equation has been transformed into an integral equation for the spectrum, but a general solution has not yet been obtained for either the spectrum or correlation.

I. INTRODUCTION

IN a recent paper Chandrasekhar[1] has presented a new theory of isotropic, homogeneous turbulence in a steady state. The basic innovations in the new theory are (1) the consideration of velocity correlations not only at two different points but also at *two different times* and (2) the hypothesis of a statistical relationship between the second-order and fourth-order correlation tensors.

Starting with the hydrodynamic equation of motion and the equation of continuity, Chandrasekhar derived his fundamental equation for $f(r,t)$, the longitudinal correlation function[2]:

$$\frac{\partial}{\partial r}\left(\frac{\partial^2}{\partial t^2}-\nu^2 D_5^2\right)f=f\frac{\partial}{\partial r}D_5 f. \tag{1}$$

Here D_5, the Laplacian operator in five-dimensional space, may be written

$$D_5=\frac{\partial^2}{\partial r^2}+\frac{4}{r}\frac{\partial}{\partial r}. \tag{2}$$

In these equations r $(=|\mathbf{r}|)$ is the separation of the two points measured in units of some arbitrary length l; t is the time interval between the measurements of velocity at the two points in the unit of $l/\langle(u_1^2)\rangle_{Av}^{\frac{1}{2}}$, where u_1 is the velocity component in the direction of $\mathbf{r}$ (say, along the x-axis); ν is the viscosity in units of $l\langle(u_1^2)\rangle_{Av}^{\frac{1}{2}}$. The longitudinal correlation function is defined in terms of velocity components by

$$f(r,t)=\frac{\langle u_1(r_0,t_0)u_1(r_0+r,\,t_0+t)\rangle_{Av}}{\langle u_1^2\rangle_{Av}}. \tag{3}$$

We define ϕ by

$$f=1-\phi; \tag{4}$$

then for small values of t and r (such that f does not

* The research reported in this paper was supported in part by the Geophysics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command, under contracts with the University of Chicago.

[1] S. Chandrasekhar, Proc. Roy. Soc. (London) **A229**, 1 (1955). Hereafter this article will be referred to as "Paper I."

[2] Paper I, Eq. (46).

depart greatly from unity) Eq. (1) may be written

$$\partial^3\phi/\partial t^3=(\nu^2D_6^3+D_6)\phi. \tag{5}$$

For the case of infinite Reynolds number (zero viscosity), Chandrasekhar has solved Eq. (5) and obtained[3]

$$f(r,t)=1-\left(\frac{r}{r_0}\right)^\alpha\psi\left(\frac{t}{r}\right), \quad (r\ll r_0; t\ll r_0) \tag{6}$$

in which[4]

$$\psi(x)=\frac{1}{2(\alpha+2)}\{|x-1|^{\alpha+2}(\alpha+2+x) +|x+1|^{\alpha+2}(\alpha+2-x)\}. \tag{7}$$

Here $x=t/r$, and in Eq. (6) r_0 is a constant of the order of the dimensions of the system in units of l.[5] In these equations α is a positive constant but is not otherwise determined. If, however, one chooses the Kolmogoroff spectrum as an initial condition, then $\alpha=\frac{2}{3}$.

In view of the interest often attached to turbulence spectra, an attempt is made in this paper to transform the correlation function of Eqs. (6) and (7) to wave-number space. In the next section the time-dependent spectrum is defined and then derived from Chandrasekhar's correlation function by means of a Fourier transform. In Sec. III, the differential equation (5) is transformed to Fourier space and is solved for the dependence of the spectrum on time when finite viscosity is considered. In Sec. IV, we transform the general, nonlinear equation (1) to wave-number space.

II. SPECTRUM FOR ZERO VISCOSITY

By analogy with the conventional definition of the turbulence spectrum, $F(k)$, we define for the time-dependent case,

$$F(k,t)=c_1k^2\{\mathbf{V}_k(t_0)\cdot\mathbf{V}_k(t_0+t)\}_{\text{av}}, \tag{8}$$

where $\mathbf{V}_k(t)$ represents the velocity at time t for an eddy of wave-number k. Throughout this paper c_i represents proportionality constants and averages are taken over all space. Note that with the definition (8), $F(k,t)$ is not necessarily always positive, since in general it is a time correlation of eddy velocities and becomes an energy spectrum only for $t=0$. The spectrum is related to the correlation function, again analogously to the conventional case,[6] by

$$f(r,t)=3\int_0^\infty\frac{F(k,t)}{k^3r^3}(\sin kr-kr\cos kr)dk\Big/\int_0^\infty F(k,0)dk. \tag{9}$$

There should be no confusion between the time-dependent spectrum for steady-state turbulence considered in this paper and the ordinary spectrum in decaying turbulence. The latter changes with *absolute* time, whereas only *relative* time is implied in the present theory. In general, for small values of t these spectra should be of comparable accuracy to the well-known Kolmogoroff spectrum (adopted herein for $t=0$) and should be applicable in such time-dependent problems as ionospheric scattering of electromagnetic radiation.[7]

In order to relate the spectrum and correlation function, we wish to express Eq. (9) in terms of a simple sine or cosine transform. Hence by successive differentiations we find

$$h(r,t)\equiv-\frac{1}{r}\frac{d}{dr}\left[\frac{1}{r}\frac{d}{dr}\left(r^4\frac{df}{dr}\right)\right]=c_2\int_0^\infty kF(k,t)\sin krdk. \tag{10}$$

It is convenient to work in the variables

$$\eta=kt, \quad y=1/x=r/t. \tag{11}$$

Then, if we define

$$j(y)\equiv t^{1-\alpha}h(r,t), \tag{12}$$

where it is assumed that the correlation may be written in the form (6), the sine transform of Eq. (10) is, for $\alpha=\frac{2}{3}$,

$$F(k,t)=\frac{c_3\eta^{\frac{1}{2}}}{k^{5/3}}\int_0^\infty j(y)\sin\eta ydy. \tag{13}$$

By means of Eqs. (12), (10), and (6), j is related to $\psi(x=1/y)$, which is given by (7). A straightforward evaluation gives two limiting series:

$$j_0(y)=-\frac{110}{81}y\left[1+\frac{(2-\frac{2}{3})(3-\frac{2}{3})}{3!}y^2 +\frac{(2-\frac{2}{3})(3-\frac{2}{3})(4-\frac{2}{3})(5-\frac{2}{3})}{5!}y^4+\cdots\right], \quad 0<y<1, \tag{14}$$

and

$$j_\infty(y)=+\frac{110}{27}y^{-1}\left[1+\frac{(1-\frac{2}{3})(2-\frac{2}{3})}{2!}y^{-2} +\frac{(1-\frac{2}{3})(2-\frac{2}{3})(3-\frac{2}{3})(4-\frac{2}{3})}{4!}y^{-4}+\cdots\right], \quad 1<y<\infty. \tag{15}$$

It will be noticed that neither of the above series converges as $y\to1$, so that $j_0\to-\infty$ and $j_\infty\to+\infty$. This discontinuity arises in the highest derivative considered in Eq. (10), and its presence may be traced

[3] Paper I, Eq. (63).
[4] Paper I, Eq. (81).
[5] In Paper I, r_0 is taken as unity and it is specified instead that the unit of length is equal to the diameter of the largest eddy present; r_0 is retained here to emphasize the uncertainty in the exact value of this constant.
[6] W. Heisenberg, Z. Physik 124, 628 (1948), Eq. (50).
[7] E.g., see R. A. Silverman and M. Balser, Phys. Rev. 96, 560 (1954).

back to the fact that for infinite Reynolds number Eq. (5) is a wave equation (in a five-dimensional space) for unit velocity. Thus physically the discontinuity in $j(y)$ at $y=1$ (or $r=t$) reflects a singularity in the correlation function at $r=t=0$. This singularity apparently arises because we have incorrectly assumed the Kolmogoroff correlation $[1-(r/r_0)^{\frac{2}{3}}]$ to hold strictly for $t=0$. And, indeed, for $a>1$ in Eq. (6) there is no discontinuity. However, the integral (13) is finite in any event and we might expect that the Kolmogoroff approximation near the origin should not have any profound effect on the spectrum, except perhaps at large wave numbers.

Substituting Eqs. (15) and (14) into (13), we obtain an expression for $F(k,t)$ involving several infinite series in powers of $\eta=kt$. If sufficient terms are carried, however, it may be shown that most of the coefficients cancel identically, with the remaining coefficients reducing to the simple expression,

$$F(k,t)=c_4k^{-5/3}\cos kt. \tag{16}$$

This equation may be checked relatively easily by taking the inverse transform as given by Eq. (9). Thus we find

$$\lim_{kr\to 0}\frac{\frac{3}{r^3}\int_{k_0}^{\infty}k^{-14/3}\cos kt(\sin kr-kr\cos kr)dk}{\int_{k_0}^{\infty}k^{-5/3}dk}$$

$$=\lim_{k_0\to 0}\left\{1-\left[\frac{27}{55}\Gamma(\tfrac{1}{3})\left(\sin\frac{\pi}{6}\right)k_0^{\frac{1}{3}}\right]r^{\frac{1}{3}}\psi\left(\frac{t}{r}\right)\right\}, \tag{17}$$

where ψ is given by Eq. (7). Comparing (17) with (6), we see that the factor in square brackets in (17) may be identified with $r_0^{-\frac{1}{3}}$, where r_0 (and k_0) refer to the largest dimensions (and eddies) in the system.

III. APPROXIMATE SPECTRUM FOR FINITE VISCOSITY

The remarkable simplification achieved in the above equations for the time-dependent spectrum suggests that the basic differential equation might also be greatly simplified if transformed to wave-number space. Therefore, writing Eq. (5) [which is the linearized form of the general equation (1)] in terms of $F(k,t)$ by means of Eq. (9) (where $f=1-\phi$), we readily obtain

$$\partial^2F/\partial t^2=(\nu^2k^2-1)k^2F. \tag{18}$$

The solution to (18) may be written:

$$F(k,t)=A\cos[(1-\nu^2k^2)^{\frac{1}{2}}kt]+B\sin[(1-\nu^2k^2)^{\frac{1}{2}}kt],\quad k<1/\nu \tag{19}$$

and

$$F(k,t)=C\cosh[(\nu^2k^2-1)^{\frac{1}{2}}kt]+D\sinh[(\nu^2k^2-1)^{\frac{1}{2}}kt],\quad k>1/\nu. \tag{20}$$

In stationary turbulence there is, by definition, no "decay" in the usual sense; the time t in these equations represents the interval of time between two events. Hence the spectrum as we have defined it in Eq. (18) cannot depend on whether t is positive or negative. Therefore, in (19) and (20) we must set $B=D=0$. Moreover, we must have $C=0$, since the spectrum must stay bounded as $t\to\infty$. Hence we are left with[8]

$$\left.\begin{aligned}F(k,t)\equiv F_\nu(k,0)\cos[(1-\nu^2k^2)^{\frac{1}{2}}kt],\quad & k<1/\nu,\\ =0,\quad & k>1/\nu.\end{aligned}\right\} \tag{21}$$

This cutoff in the spectrum for $k>1/\nu$ apparently indicates that the turbulence has been completely overcome by viscosity and that for sufficiently small elements of the fluid only laminar flow exists. Heisenberg[6] has pointed out that there may be a turbulence cutoff at high wave numbers and that consequently his k^{-7} law might not hold for indefinitely small eddies. This idea has been elaborated upon by Batchelor and Townsend.[9]

It is not possible at present to specify the nature of $F_\nu(k,0)$, which appears as an integration constant in Eq. (21). Probably Heisenberg's[6] spectrum represents the closest approach to the energy distribution presently available; however, it can only be an approximation, for this initial energy spectrum must also have a cutoff at $k=1/\nu$. That is, from (21) we have

$$F(1/\nu,t)=F_\nu(1/\nu,0)=0. \tag{22}$$

The cosine factor in (21) seems to indicate an oscillation of the eddy velocities between the limits $\pm|\mathbf{V}_k(t=0)|$. However, since we are restricted to stationary turbulence, the *absolute values* of the velocities cannot be a function of time. That is, the fluctuation of $F(k,t)$ with time merely reflects a constant rotation of the eddy $\mathbf{k}$ in space, with a period of

$$T=2\pi/[(1-\nu^2k^2)^{\frac{1}{2}}k]. \tag{23}$$

For $k=0$ the rotation period is infinite and decreases to a minimum of $T_{\min}=4\pi\nu$ for eddies with $k=1/\sqrt{2}\nu$. For smaller eddies the viscosity hinders the rotation more and more until finally at $k=1/\nu$ there is no turbulent motion whatever. Thus the cosine term is interpreted as nothing more than a consequence of isotropy in three-dimensional turbulence.

With this picture the eddies are perfectly stable and retain their identity indefinitely. Apparently, quantitative information on the finite lifetimes of eddies [as well as further information on the initial spectrum, $F_\nu(k,0)$] must await a solution to the more exact

[8] One's first inclination is to write $F(k,t)=C\exp[-(\nu^2k^2-1)^{\frac{1}{2}}kt]$ for $k>1/\nu$; however, this solution is unacceptable for the reasons stated above. We are indebted to Professor Chandrasekhar for drawing our attention to this point and for demonstrating that the spectrum must vanish instead for $k>1/\nu$.

[9] G. K. Batchelor and A. A. Townsend, Proc. Roy. Soc. (London) **A199**, 238 (1949).

differential equation (1) or its equivalent in wave number space, which is derived in the next section.

IV. GENERAL EQUATION FOR THE SPECTRUM

In this section, we wish to express the general differential equation (1) in terms of the turbulence spectrum, $F(k,t)$. Equation (9) may be written in terms of Bessel functions, J_s, as follows[8]:

$$f(r,t)=c_5\int_0^\infty \frac{F(k',t)}{(k'r)^{\frac{1}{2}}}J_{\frac{1}{2}}(k'r)dk', \tag{24}$$

where

$$c_5=3(\pi/2)^{\frac{1}{2}}\Big/\int_0^\infty F(k,0)dk. \tag{25}$$

Combining (24) with a recurrence relation for Bessel functions,

$$\frac{d}{dz}[z^{-s}J_s(z)]=-z^{-s}J_{s+1}(z), \tag{26}$$

we obtain for the left-hand side of Eq. (1),

$$\frac{\partial}{\partial r}\left(\frac{\partial^2}{\partial t^2}-\nu^2D_5^2\right)f=-\frac{c_5}{r^2}\int_0^\infty\left[\frac{F}{k'}\left(\frac{\partial^2}{\partial t^2}-\nu^2k'^4\right)\right]\times J_{\frac{1}{2}}(k'r)(k'r)^{\frac{1}{2}}dk'. \tag{27}$$

The right-hand side of Eq. (1) similarly transforms to

$$f\left(\frac{\partial}{\partial r}D_5f\right)=-c_5^2\int_0^\infty F(k',t)\frac{J_{\frac{1}{2}}(k'r)}{(k'r)^{\frac{1}{2}}}dk' \times\int_0^\infty\frac{k''^3F(k'',t)}{r}\frac{d}{dk''}\left[\frac{J_{\frac{1}{2}}(k''r)}{(k''r)^{\frac{1}{2}}}\right]dk''. \tag{28}$$

Integrating the second integral in (28) by parts and setting Eq. (27) equal to (28), we have, for a spectrum that vanishes sufficiently rapidly as $k\to\infty$,

$$\int_0^\infty\left[\frac{1}{k'}\left(\frac{\partial^2F}{\partial t^2}-\nu^2k'^4F\right)\right]J_{\frac{1}{2}}(k'r)(k'r)^{\frac{1}{2}}dk' =-\frac{c_5}{r^2}\int_0^\infty\int_0^\infty F(k't)\frac{\partial}{\partial k''}[k''^3F(k'',t)] \times\frac{J_{\frac{1}{2}}(k'r)J_{\frac{1}{2}}(k''r)}{(k'k'')^{\frac{1}{2}}}dk'dk''. \tag{29}$$

Next we multiply both sides of Eq. (29) by $(kr)^{\frac{1}{2}}J_{\frac{1}{2}}(kr)$ and integrate over r from zero to infinity. Then applying Hankel's inversion theorem[10] to the left-hand side of (29), we find, with the aid of Eq. (25),

$$\left(\frac{\partial^2}{\partial t^2}-\nu^2k^4\right)F(k,t) =-\frac{\frac{1}{k}\int_0^\infty\int_0^\infty K(k,k',k'')F(k't)\frac{\partial}{\partial k''}[k''^3F(k'',t)]dk'dk''}{\int_0^\infty F(k,0)dk}, \tag{30}$$

where

$$K(k,k',k'')=\frac{3(\pi/2)^{\frac{1}{2}}k^{\frac{1}{2}}}{(k'k'')^{\frac{1}{2}}}\int_0^\infty\frac{J_{\frac{1}{2}}(kr)J_{\frac{1}{2}}(k'r)}{r^{\frac{1}{2}}} \times J_{\frac{1}{2}}(k''r)dr. \tag{31}$$

The kernel K has been evaluated[11] as follows:

$$K(k,k',k'')=\begin{cases}0, & (k'-k'')^2>k^2,\\ 1, & (k'+k'')^2<k^2,\\ \frac{1}{4}(\cos\theta-1)^2(\cos\theta+2), & \end{cases} \tag{32}$$

otherwise, where

$$\cos\theta=(k'^2+k''^2-k^2)/2k'k''. \tag{33}$$

One additional integration by parts of Eq. (30) gives

$$\left(\frac{\partial^2}{\partial t^2}-\nu^2k^4\right)F(k,t) =\frac{\frac{1}{k}\int_0^\infty\int_0^\infty k''^3F(k't)F(k''t)\frac{\partial K(k,k',k'')}{\partial k''}dk'dk''}{\int_0^\infty F(k,0)dk}. \tag{34}$$

Although $\partial K/\partial k''$ is a discontinuous function, Eq. (34) may be more amenable to numerical computations than Eq. (30).

Equation (30) or (34), with K as given by (32), is the equivalent in wave-number space of Chandrasekhar's general equation (1). The relatively simple form of Eq. (32) for the kernel, K, suggests that (34) may prove useful for investigating the spectrum of turbulence in Chandrasekhar's new theory.

V. ACKNOWLEDGMENTS

Professor Chandrasekhar kindly suggested that we undertake this extension to his recent work. We are also deeply grateful to him for access to his paper prior to publication and for numerous illuminating discussions on turbulence theory.

[10] I. N. Sneddon, *Fourier Transforms* (McGraw-Hill Book Company, Inc., New York, 1951), p. 48.

[11] G. N. Watson, *A Treatise on Bessel Functions* (University Press, Cambridge, 1944), second edition, p. 411, Eq. (2).

Solution of Problems Involving the Hydromagnetic Flow of Compressible Ionized Fluids

R. E. LOUGHHEAD
Division of Physics, National Standards Laboratory, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia
(Received March 29, 1955)

The characteristic forms of the hydromagnetic equations for a compressible fluid are examined from the viewpoint of obtaining by the method of finite differences numerical solutions for continuous initial value problems involving unidimensional motion of the fluid.

INTRODUCTION

THE equations of motion of an ionized fluid are complicated in form and are frequently simplified by assuming that (a) the electrical conductivity is infinite, and (b) the fluid is incompressible. The first of these assumptions leads to the well-known concept of the 'freezing-in' of the lines of magnetic force and has been discussed in detail elsewhere, e.g., by Cowling.[1] The second assumption is admissible provided that the fluid velocities encountered are small compared with the velocity of sound.[2] However, this condition is not always satisfied in astrophysical problems, where the gas velocity may be comparable with or even exceed the local velocity of sound.

Hence in order to study many aspects of the hydromagnetic behavior of ionized fluids account must be taken of the compressibility of the medium. At the same time, some assumption must be made concerning the thermal conditions in the gas, and it is convenient to adopt the usual postulate of adiabatic conditions in the gas.

Retaining the assumption (a) it is then found that the equations of motion of an ionized fluid in a magnetic field yield a set of simultaneous partial differential equations which are nonlinear and therefore highly intractable from the viewpoint of obtaining analytical solutions. Some progress has been made by restricting consideration to small perturbations about some given state, when the equations may be linearized by the neglect of second order terms. However, for many problems account must be taken of the intrinsic nonlinearity of the equations since the linearized forms are quite inapplicable.

In particular, this is the case for initial value problems in which it is desired to predict the behavior of an ionized gas given that it is in some specified state at an initial time $t=0$. Problems of this type could, for instance, have an important bearing on the so-called 'inverse-skin' effect in the solar atmosphere, which may be involved in the formation of solar prominences.[3]

Similar nonlinear problems are encountered in other branches of physics, particularly in aerodynamics where there are not the complications arising from the presence of a magnetic field in an ionized gas.[4] In aerodynamics considerable progress has been made in obtaining numerical solutions for unidimensional gas flow problems using the method of characteristics. This involves forming linear combinations of the original equations to form an equivalent set of relations, in each of which all the variables are differentiated in the same direction along a common curve. The set of curves so arising are called the 'characteristics' of the original equations, and their gradients the 'characteristic directions.'

In aerodynamics the application of such methods has largely been restricted to problems involving only a pair of characteristic equations in two independent variables. However Courant, Isaacson, and Rees[5] have formulated a finite difference scheme for the numerical solution of continuous initial value problems described by a set of any (finite) number of characteristic equations in two independent variables. An examination of the characteristic forms of the hydromagnetic equations for a compressible, ionized fluid indicates that a finite difference scheme along the lines proposed by Courant *et al.* could be used with considerable advantage to obtain numerical solutions for continuous initial value problems involving unidimensional motion of the fluid. An outline of the procedure is given in this paper. The numerical work entailed is extremely lengthy and no attempt is made here to carry through the full solution for any specific problem.

APPLICATION OF THE METHOD OF CHARACTERISTICS TO A PROBLEM IN UNIDIMENSIONAL FLOW

Consider the unidimensional motion of a fluid in which the magnetic field is everywhere at right angles to the direction of motion, a simple case whose importance has been pointed out by Alfven.[3] With the assumptions of infinite electrical conductivity and adiabatic conditions following the motion of the gas the hydromagnetic equations for a compressible ionized fluid may be written in the well-known forms (e.g.,

[1] T. G. Cowling, *The Sun*, edited by G. P. Kuiper (University of Chicago Press, Chicago, 1952), Chap. 8.
[2] L. Howarth, *Modern Developments in Fluid Dynamics* (Oxford University Press, London, 1953), Vol. 1, Chap. 1.
[3] H. Alfven, *Cosmical Electrodynamics* (Oxford University Press, London, 1950).
[4] R. Courant and K. O. Friedrich, *Supersonic Flow and Shock Waves* (Interscience Publishers, Inc., New York, 1948).
[5] Courant, Isaacson, and Rees, Com. Pure Appl. Maths. **5**, 243 (1952).

Cowling):

$$\partial \mathbf{H}/\partial t = \operatorname{curl}(\mathbf{v}\times\mathbf{H}), \tag{1}$$

$$4\pi\rho D\mathbf{v}/Dt = (\operatorname{curl}\mathbf{H})\times\mathbf{H} - 4\pi \operatorname{grad} p, \tag{2}$$

$$D\rho/Dt + \rho \operatorname{div}\mathbf{v} = 0, \tag{3}$$

$$Dp/Dt + \gamma p \operatorname{div}\mathbf{v} = 0, \tag{4}$$

where the operator

$$D/Dt \equiv \partial/\partial t + (\mathbf{v}\cdot\operatorname{grad})$$

denotes differentiation following the motion of the fluid, and the symbols $\mathbf{H}$, $\mathbf{v}$, p, ρ, and γ denote respectively magnetic field, fluid velocity, scalar pressure, mass density, and ratio of the specific heats of the fluid. The units employed are Gaussian.

It is convenient to introduce cartesian axes Ox, y, z such that the magnetic field is directed in the y-direction and the motion of the fluid occurs parallel to the x-axis. Then, writing

$$H = H_y \quad \text{and} \quad v = v_x,$$

the basic equations (1)$\cdots$(4) yield in the case of unidimensional flow depending on the single spatial coordinate x the relations

$$v\frac{\partial H}{\partial x} + H\frac{\partial v}{\partial x} + \frac{\partial H}{\partial t} = 0, \tag{5}$$

$$H\frac{\partial H}{\partial x} + 4\pi\rho v\frac{\partial v}{\partial x} + 4\pi\rho\frac{\partial v}{\partial t} + 4\pi\frac{\partial p}{\partial x} = 0, \tag{6}$$

$$\rho\frac{\partial v}{\partial x} + v\frac{\partial \rho}{\partial x} + \frac{\partial \rho}{\partial t} = 0, \tag{7}$$

$$\gamma p\frac{\partial v}{\partial x} + v\frac{\partial p}{\partial x} + \frac{\partial p}{\partial t} = 0. \tag{8}$$

These equations are hyperbolic in type and their characteristic forms may be conveniently expressed by introducing

$$V = \left(\frac{H^2}{4\pi\rho} + \frac{\gamma p}{\rho}\right)^{\frac{1}{2}} \tag{9}$$

to denote the local value of the hydromagnetic wave velocity. Then the equations may be transformed into the following characteristic set (see Appendix):

$$\rho\left\{\frac{\partial H}{\partial t} + v\frac{\partial H}{\partial x}\right\} - H\left\{\frac{\partial \rho}{\partial t} + v\frac{\partial \rho}{\partial x}\right\} = 0, \tag{10}$$

$$\rho\left\{\frac{\partial p}{\partial t} + v\frac{\partial p}{\partial x}\right\} - \gamma p\left\{\frac{\partial \rho}{\partial t} + v\frac{\partial \rho}{\partial x}\right\} = 0, \tag{11}$$

$$H\left\{\frac{\partial H}{\partial t} + (v+V)\frac{\partial H}{\partial x}\right\} + 4\pi\rho V\left\{\frac{\partial v}{\partial t} + (v+V)\frac{\partial v}{\partial x}\right\} + 4\pi\left\{\frac{\partial p}{\partial t} + (v+V)\frac{\partial p}{\partial x}\right\} = 0, \tag{12}$$

$$H\left\{\frac{\partial H}{\partial t} + (v-V)\frac{\partial H}{\partial x}\right\} - 4\pi\rho V\left\{\frac{\partial v}{\partial t} + (v-V)\frac{\partial v}{\partial x}\right\} + 4\pi\left\{\frac{\partial p}{\partial t} + (v-V)\frac{\partial p}{\partial x}\right\} = 0. \tag{13}$$

In each of these equations the variables H, v, p, and ρ are all differentiated along a common curve in the $x-t$ plane whose slope is given by

$$dx/dt = (v, v+V, v-V) \tag{14}$$

respectively. When the magnetic field vanishes Eqs. (10)–(13) reduce to the characteristic forms of the equations describing unidimensional adiabatic flow of aerodynamical fluids. In this case the hydromagnetic wave velocity V is replaced by the velocity of sound in the fluid.

Consider now a rectangular mesh in the $x-t$ plane whose (k,l)th point is defined by

$$x_k = k\Delta x, \quad t_l = l\Delta t, \tag{15}$$

where

$$r = \Delta t/\Delta x \tag{16}$$

si the mesh ratio. It is assumed that the slopes of the characteristics are bounded throughout the region of integration. If

$$|dx/dt|_{\max} < M \tag{17}$$

throughout this region, the mesh ratio r must be chosen so as to satisfy the Courant-Friedrich-Lewy condition

$$r < 1/M \tag{18}$$

to insure the stability of the finite difference scheme.

The technique is now to replace derivatives along characteristics by finite differences, the choice of difference quotients (forward or backward) being made so as to preserve the domain of dependence.[5] If $H_{k,l}$, $v_{k,l}$, $p_{k,l}$, and $\rho_{k,l}$ denote the values of the solutions of the difference equations at the net point (x_k, t_l), the characteristic equations are replaced by the following set of difference equations:

$$\begin{aligned}\rho_{k,l}H_{k,l+1} - H_{k,l}\rho_{k,l+1} &= \rho_{k,l}\{(1\pm rv_{k,l})H_{k,l} \mp rv_{k,l}H_{k\pm1,l}\} \\ &\quad - H_{k,l}\{(1\pm rv_{k,l})\rho_{k,l} \mp rv_{k,l}\rho_{k\pm1,l}\},\end{aligned} \tag{19}$$

$$\rho_{k,i}p_{k,i+1}-\gamma p_{k,i}\rho_{k,i+1}$$
$$=\rho_{k,i}\{(1\pm\tau v_{k,i})p_{k,i}\mp\tau v_{k,i}p_{k\pm1,i}\}$$
$$-\gamma p_{k,i}\{(1\pm\tau v_{k,i})\rho_{k,i}\mp\tau v_{k,i}\rho_{k\pm1,i}\}. \quad (20)$$

$$H_{k,i}H_{k,i+1}+4\pi(\rho V)_{k,i}v_{k,i+1}+4\pi p_{k,i+1}$$
$$=H_{k,i}\{[1\pm\tau(v+V)_{k,i}]H_{k,i}\mp\tau(v+V)_{k,i}H_{k\pm1,i}\}$$
$$+4\pi(\rho V)_{k,i}\{[(1\pm\tau(v+V)_{k,i}]v_{k,i}\mp\tau(v+V)_{k,i}v_{k\pm1,i}\}$$
$$+4\pi\{[1\pm\tau(v+V)_{k,i}]p_{k,i}\mp\tau(v+V)_{k,i}p_{k\pm1,i}\}. \quad (21)$$

$$H_{k,i}H_{k,i+1}-4\pi(\rho V)_{k,i}v_{k,i+1}+4\pi p_{k,i+1}$$
$$=H_{k,i}\{[1\pm\tau(v-V)_{k,i}]H_{k,i}\mp\tau(v-V)_{k,i}H_{k\pm1,i}\}$$
$$-4\pi(\rho V)_{k,i}\{[1\pm\tau(v-V)_{k,i}]v_{k,i}\mp\tau(v-V)_{k,i}v_{k\pm1,i}\}$$
$$+4\pi\{[1\pm\tau(v-V)_{k,i}]p_{k,i}\mp\tau(v-V)_{k,i}p_{k\pm1,i}\}. \quad (22)$$

In each of Eqs. (19)–(22) the upper or lower alternative signs are taken according as the respective characteristic slope is negative or positive.

The difference equations possess nontrivial solutions only if the quantity

$$8\pi\rho^2V\{H^2+4\pi\gamma p\}$$

does not vanish in the region of integration. This condition is clearly satisfied.

The solution of these equations, entailing the calculation of a new set of characteristics at each point of the mesh, is extremely laborious and no attempt is made here to carry out the numerical work involved.

However, it is of interest to note an example of the type of problem which could be investigated by an appropriate choice of initial conditions. Consider the case of a magnetic field which at time $t=0$ is largely concentrated within a given region of an ionized gas of infinite extent. For definiteness the spatial distribution of the field at $t=0$ may be assumed to follow the Gaussian law

$$H(x,0)=H_0\exp(-\alpha x^2),$$

where H_0 and α are constants. Suppose, also, that the mass velocity at this instant vanishes everywhere, while the pressure and mass density have the uniform values p_0 and ρ_0. The initial conditions of the problem would thus take the forms

$$\begin{aligned} H_{k,0}&=H_0\exp(-\alpha x_k^2),\\ v_{k,0}&=0,\\ p_{k,0}&=p_0,\\ \rho_{k,0}&=\rho_0, \end{aligned} \quad (23)$$

where k is any integer, positive or negative. By assuming physically sensible values for the constants H_0, p_0, ρ_0, and α, the numerical solution could then be stepped off from the initial conditions (23) with the aid of the difference equations given above.

ACKNOWLEDGMENTS

The author is indebted to Professor V. A. Bailey and to Dr. R. G. Giovanelli and Dr. J. A. Roberts for the benefit of helpful discussions.

APPENDIX

The characteristic equations (10)$\cdots$(13) may be obtained by forming appropriate linear combinations of the basic equations (5)$\cdots$(8). To do this it is convenient to introduce the upper suffix notation:

$$u^i=(H,v,p,\rho);\quad i=1,2,3,4,$$

and to use the subscripts to denote partial differentiation with respect to the variables x and t. In this notation the fundamental equations read

$$a^{ij}u_x{}^j+b^{ij}u_t{}^j=0, \quad (24)$$

where the repeated indices are to be summed from 1 to 4 and the coefficients a and b are displayed as the matrix arrays:

$$|a^{ij}|=\begin{bmatrix} v & H & 0 & 0\\ H & 4\pi\rho v & 4\pi & 0\\ 0 & \rho & 0 & v\\ 0 & \gamma p & v & 0 \end{bmatrix};$$
$$|b^{ij}|=\begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & 4\pi\rho & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (25)$$

The aim of the calculation is to replace Eqs. (24) by an equivalent set:

$$\lambda^{\nu i}b^{ij}(C^\nu u_x{}^j+u_t{}^j)=0;\quad \nu=1,2,3,4 \quad (26)$$

in each of which the variables u are differentiated in a common direction

$$du^i/dt=C^\nu u_x{}^i+u_t{}^i,$$

along a curve $x=x(t)$ for which $dx/dt=C^\nu$ (the νth characteristic direction).

To effect this transformation, multipliers λ^i are sought such that

$$\lambda^i a^{ij}=\lambda^i C b^{ij}, \quad (27)$$

and the condition for the existence of λ^i, not all zero, is that C satisfy the characteristic equation

$$|a^{ij}-Cb^{ij}|=0. \quad (28)$$

This equation has the four roots

$$C^\nu=(v,v,v+V,v-V);\quad \nu=1,2,3,4 \quad (29)$$

which are all real, thus verifying that Eqs. (24) are of hyperbolic type.

Corresponding to the four roots C^ν four linearly independent sets of multipliers $\lambda^{\nu 1},\cdots,\lambda^{\nu 4}$ ($\nu=1,2,3,4$) are required satisfying the relations (27) when $C=C^\nu$. In the case of a nonrepeated root the corresponding multipliers are proportional to the cofactors of the

elements of any column of the determinant (28). By using the cofactors of the elements of the fourth column it is found that

$$\frac{\lambda^{\nu 1}}{H}=\frac{-\lambda^{\nu 3}}{v-C^\nu}=\frac{\lambda^{\nu 3}}{4\pi\{(v-C^\nu)^2-V^2\}}=\frac{\lambda^{\nu 4}}{4\pi},$$

so that the characteristic equation corresponding to a nonrepeated root C^ν takes the form:

$$H\left\{\frac{\partial H}{\partial t}+C^\nu\frac{\partial H}{\partial x}\right\}-4\pi\rho(v-C^\nu)\left\{\frac{\partial v}{\partial t}+C^\nu\frac{\partial v}{\partial x}\right\}+4\pi\left\{\frac{\partial p}{\partial t}+C^\nu\frac{\partial p}{\partial x}\right\}=0. \quad (30)$$

By substituting in turn $C^3=v+V$ and $C^4=v-V$ in (30) the characteristic Eqs. (12) and (13) are immediately obtained.

In the case of the repeated root $C^1=C^2=v$, two linearly independent sets of multipliers λ^{1i} and λ^{2i} are required such that

$$\lambda^{\nu i}(a^{ij}-vb^{ij})=0,$$

if $\nu=1$ or 2. By noting that the matrix $a^{ij}-vb^{ij}$ is of rank 2 and must therefore possess two linearly independent left-hand factors of zero, it is found that

$$\lambda^{1i}=(\rho,0,-H,0),$$
$$\lambda^{2i}=(0,0,\gamma p,-\rho)$$

are two such linearly independent sets of multipliers. Hence the remaining characteristic equations are obtained in the forms:

$$\rho\left\{\frac{\partial H}{\partial t}+v\frac{\partial H}{\partial x}\right\}-H\left\{\frac{\partial \rho}{\partial t}+v\frac{\partial \rho}{\partial x}\right\}=0,$$
$$\rho\left\{\frac{\partial p}{\partial t}+v\frac{\partial p}{\partial x}\right\}-\gamma p\left\{\frac{\partial \rho}{\partial t}+v\frac{\partial \rho}{\partial x}\right\}=0, \quad (31)$$

which agree with Eqs. (10) and (11).

The four sets of multipliers must be linearly independent, and this is ensured by the nonvanishing of the determinant

$$|\lambda^{\nu i}|=-2\rho V(4\pi\gamma p+H^2)$$

under all physically relevant conditions.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Paramagnetic Resonance at Low Fields*

STEWART BECKER†
Physics Department, Agricultural and Mechanical College of Texas, College Station, Texas
(Received March 7, 1955; revised manuscript received May 31, 1955)

Paramagnetic resonance in 2,2-diphenyl-1-picrylhydrazyl and Wurster's Blue at very low values of applied magnetic fields has been investigated. The disappearance of the maximum of the absorption curve below frequencies of 1.3 megacycles has been substantiated. A twin-T resonance spectrometer is described briefly. Line broadening with dilution of diphenylpicrylhydrazyl in benzene was investigated at 35 gauss and the second and fourth moments of the absorption curves calculated. These data show that the fourth moment increases more than the second moment as dilution is increased.

1. INTRODUCTION

A CONSIDERABLE amount of work has been done in recent years on paramagnetic resonance in 2,2-diphenyl-1-picrylhydrazyl at fields ranging from 10 to 10 000 gauss,[1–4] and more recently at fields as low as 0.45 gauss.[5] This paper presents information gained from a series of measurements made on crystalline samples of both diphenylpicrylhydrazyl and Wurster's Blue in fields ranging from 0.24 gauss to 35 gauss, and on line broadening with dilution in benzene of diphenylpicrylhydrazyl at a field of 35 gauss. The experimental results are then compared with theory as developed for the low field case by Garstens,[6] and with the theory of exchange narrowing as developed by Van Vleck.[7]

2. APPARATUS

A resonance spectrometer using a twin-T bridge as shown in Fig. 1 was used. All measurements were made at room temperature. The samples consisted of 15 grams of polycrystalline diphenylpicrylhydrazyl, 5 grams of polycrystalline Wurster's Blue, and 3-cc liquid samples of diphenylpicrylhydrazyl in benzene.

2.1 Static Magnetic Field

The static magnetic field was produced by a solenoid wound on an aluminum tube six inches in diameter and

* Supported in part by Research Corporation.
† Present address: Physics Department, University of Arizona, Tucson, Arizona.
[1] L. S. Singer and E. G. Spencer, J. Chem. Phys. **21**, 939 (1953).
[2] Hutchison, Pastor, and Kowalsky, J. Chem. Phys. **20**, 534 (1952).
[3] A. N. Holden, J. Chem. Phys. **19**, 1319 (1951).
[4] Holden, Kittel, Merritt, and Yager, Phys. Rev. **77**, 147 (1950).
[5] Garstens, Singer, and Ryan, Phys. Rev. **96**, 53 (1954).
[6] M. A. Garstens, Phys. Rev. **93**, 1228 (1954).
[7] J. H. Van Vleck, Phys. Rev. **74**, 1168 (1948).

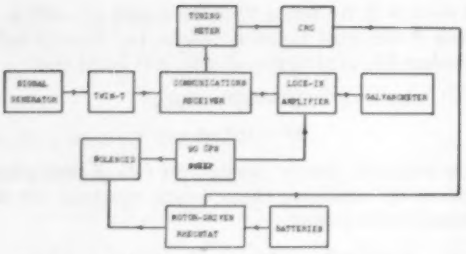


FIG. 1. Block diagram of twin-T resonance spectrometer.

three feet long. This form was threaded 52 threads per inch and wound with 26 layers of No. 26 NYCLAD copper wire. The solenoid was lined up with the earth's magnetic field by a compass and dip-needle, and one of the layers of wire was used to produce a field equal in magnitude and oppositely directed to the earth's field. This insured that the applied field could be swept to zero.

With the twin-T resonance spectrometer, it was much easier to sweep the applied magnetic field than the radio frequency, and this was done by a motor-driven rheostat. In order to insure that the zero field point appeared on the oscillograms it seemed best to actually reverse the field in the solenoid by means of the motor driven sweep. Thus the resonance curve appeared as a double peak on the oscilloscope. Actually these two peaks are the peaks of the same resonance curve, one occurring when the direction of the field in the solenoid was toward one end, and the other occurring when the field in the solenoid was toward the other end. This was of considerable value in determining that an absorption maximum was lacking at fields below 0.47 gauss as predicted by Garstens.[6]

2.2 Twin-T Bridge

A twin-T bridge as described by Sinclair[8] was used. This bridge was balanced to one part in ten million by means of vernier capacitors consisting of quarter-inch brass rods threaded 48 threads to the inch for the grounded plate, with fixed brass plates spaced approximately one-half inch from the ends of these rods serving as the other plates. Verniers capable of finer adjustment than this would have been desirable. The bridge housing was made of one-quarter inch aluminum plates and consisted of four shielded compartments. The controls were entirely orthogonal at the lower frequencies, but interaction was noted at the higher frequencies. The samples were contained in thin-walled test tubes and the bridge coils were wound directly on the test tubes to obtain the optimum filling factor.

2.3 Signal Source

A general Radio Signal Generator Type 1001 A was used for measurements below 50 megacycles, and a crystal controlled signal generator was used at 90 megacycles. Provision was made for checking frequencies against WWV.

[8] D. B. Sinclair, Proc. Inst. Radio Engrs. 28, 310 (1940).

2.4 Detection System

The detection system consisted of a Hallicrafters SX–62 communications receiver augmented by a differential type dc tuning meter. The output of this tuning meter was applied to the vertical deflection amplifier of the cathode ray oscilloscope, and a sweep voltage of approximately one-fifth of a second from the motor driven rheostat controlling the magnetic field was applied to the horizontal deflection amplifier. With proper bridge tuning this system supplied direct absorption curves.

Derivative curves of the absorption signal were obtained by a lock-in amplifier triggered by the 90 cps sine wave oscillator producing modulation of the magnetic field.

3. EXPERIMENTAL RESULTS AND DISCUSSION

A typical oscillogram showing a direct absorption curve for crystalline diphenylpicrylhydrazyl at a "steady" magnetic field in the solenoid of 0.73 gauss with a radio frequency of 2 megacycles applied to the twin-T bridge is shown in Fig. 2. For this direct absorption curve, the "steady" magnetic field in the solenoid was swept by the motor driven rheostat approximately 2.5 gauss each side of zero field, i.e., the field in the solenoid was reversed as previously explained. This method insured that the point of zero applied field was observed. It was very easy to check from the oscillograms whether the solenoid was properly lined up with the earth's magnetic field and that the correct field was applied to the bucking winding to cancel this field by checking to see that the dip in the double curve occurred at the zero current point in the

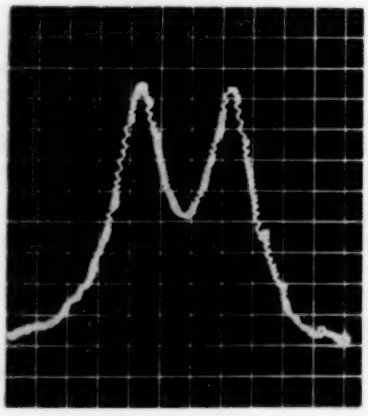


FIG. 2. Paramagnetic resonance absorption in diphenylpicryl hydrazyl at a field of 0.73 gauss with an applied radio-frequency of 2 megacycles.

solenoid. The frequency of the motor driven sweep was one-fifth cycle per second.

Figure 3 shows the direct absorption curve of diphenylpicrylhydrazyl at a "steady" magnetic field in the solenoid of 0.27 gauss with a radio-frequency of 0.6 megacycle applied to the twin-T bridge. The lack of the double peak which was evident in Fig. 2 is in accordance with Garsten's prediction[6] that no absorption maximum should be present below 1.3 megacycles. The single peak shown in Fig. 3 is the point of zero field in the solenoid.

Line widths for diphenylpicrylhydrazyl at fields ranging from 5 to 35 gauss were studied and were in substantial agreement with those recently reported by Garsten's.[5] These widths were obtained by using a 90 cps sweep approximately 0.25 gauss wide and slowly varying the steady magnetic field by a hand operated rheostat to observe the derivative of the absorption curve and record the points of maximum slope.

3.1 Liquid Samples

Solutions of diphenylpicrylhydrazyl in benzene ranging from 0.3 molar to 0.04 molar were studied and the second and fourth moments of the absorption curve calculated. Van Vleck has predicted[7] that dipolar broadening is decreased as molecular spacing is increased, and that this should result in little change in the second moment of the absorption curve, but a substantial increase in the fourth moment. In addition, nuclear hyperfine interactions become a factor in determining line width and shape as dilution is increased. At first these interactions cause line broadening beyond that predicted by dipole-dipole interactions alone, and finally they cause the smooth line to break up into several well defined lines. A plot of the second and fourth moments as dilution is increased is shown in Fig. 4. Data for these curves were obtained from enlarged oscillograms using the definition of the nth moment as

$$M_n = \int_{H=0}^{H=\infty} H^n dA,$$

where H refers to the field where the amplitude of the absorption is A, and where the resonance lines have been normalized by setting

$$\int dA = 1.$$

Since some dispersive signal combined with the absorption signal resulted in asymmetry of the absorption line, each half of the curve was analyzed separately and the average of the moments of each half was used in plotting the curves shown in Fig. 4. Although the second moment did not remain entirely constant as predicted by Van Vleck (and this variation could well be attributed to hyperfine nuclear interactions) the fourth moment showed a greater percentage increase for all solutions tested.

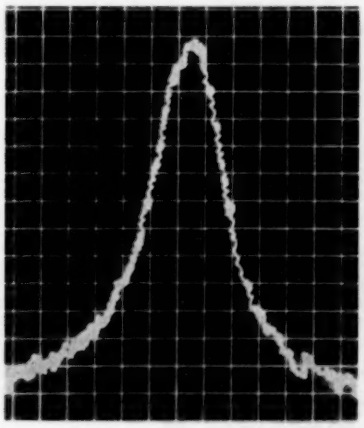


FIG. 3. Paramagnetic resonance absorption in diphenylpicryl hydrazyl at a field of 0.27 gauss with an applied radio-frequency of 0.6 megacycle.

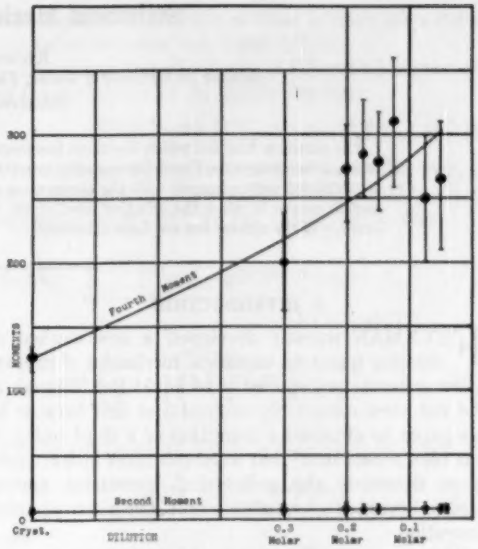


FIG. 4. Second and fourth moments; diphenyl-picryl hydrazyl.

4. ACKNOWLEDGMENTS

The author is greatly indebted to Dr. Melvin Eisner for his encouragement and many helpful suggestions during this investigation.

Statistical Mechanics of Liquid He^3

RYOICHI KIKUCHI*
Institute for the Study of Metals, The University of Chicago, Chicago, Illinois
(Received May 9, 1955)

The partition function which Feynman has proposed for liquid He^4 is extended to the case of liquid He^3, taking into account the Fermi-Dirac statistics and the nuclear spin of $\frac{1}{2}$. The nuclear magnetic susceptibility is calculated and compared with the observation of Fairbank, Ard, and Walters. Theoretical and experimental curves fit when the effective mass of the He^3 atom is taken about four times its true mass. The entropy of the system has not been calculated.

1. INTRODUCTION

FEYNMAN recently developed[1] a new method of treating quantum statistical mechanics of cooperative systems, and applied it to liquid He^4. Though it did not seem completely successful at first because in his paper he obtained a transition of a third order, it has been shown later[2] that when one takes into account more rigorously the geometrical correlation among atoms, Feynman's formulation does give a second-order transition.

This success of Feynman's approach encourages us to go one step further and examine the case of liquid He^3, which obeys Fermi-Dirac statistics, in order to make the discussion about liquid helium complete. The present paper is an attempt in this direction and discusses the problem based on the partition function and the lattice model Feynman introduced in F I.

2. ALL PARALLEL SPINS

The difference between He^4 and He^3 is not only the statistics but that the latter has a nuclear spin of magnitude $\frac{1}{2}$, whereas the former has not. In order to show the difference between the Bose-Einstein and the Fermi-Dirac statistics clearly, in this section we shall treat a case in which the effect of the spin can be neglected, namely the state of liquid He^3 in which all the nuclear spins are aligned in the same direction.

The expression of the partition function for Bose-Einstein particles Feynman derived is shown in Eq. (5) of F I and can be written in the following form:

$$Q_{BE}=N!^{-1}\sum_P\int(P\mathbf{z}_1,P\mathbf{z}_2,\cdots P\mathbf{z}_N \times|e^{-\beta H}|\mathbf{z}_1,\cdots,\mathbf{z}_N)d\mathbf{z}_1\cdots d\mathbf{z}_N, \quad (2.1)$$

in which the function in the integrand is defined by

$$(\mathbf{z}_1',\mathbf{z}_2',\cdots,\mathbf{z}_N'|e^{-\beta H}|\mathbf{z}_1,\mathbf{z}_2,\cdots,\mathbf{z}_N) = \int_{\mathbf{z}_i'}\exp\left\{-\int_0^\beta\left[\frac{m}{2\hbar^2}\sum_i\left(\frac{d\mathbf{x}_i}{du}\right)^2 + \sum_{i,j}V(\mathbf{x}_i-\mathbf{x}_j)\right]du\right\}\mathfrak{D}^N\mathbf{x}_i(u), \quad (2.2)$$

where for each trajectory the following conditions are to be satisfied:

$$\mathbf{x}_i(0)=\mathbf{z}_i \quad \text{and} \quad x_i(\beta)=z_i'. \quad (2.3)$$

P in Eq. (2.1) is a permutation of the coordinates of N particles among themselves. For the derivation of these expressions and the detail of the notations, readers are referred to the original paper, F I.

It can easily be shown that for the partition function Q_{FD} for Fermi-Dirac particles one has to bring in a factor $(-1)^P$ into Eq. (2.1), so that

$$Q_{FD}=N!^{-1}\sum_P(-1)^P\int(P\mathbf{z}_1,P\mathbf{z}_2,\cdots P\mathbf{z}_N \times|e^{-\beta H}|\mathbf{z}_1,\cdots\mathbf{z}_N)d\mathbf{z}_1\cdots d\mathbf{z}_N. \quad (2.4)$$

The integrand of Eq. (2.4) is the same function as the integrand of Eq. (2.1) and is given by Eq. (2.2), independent of the kind of statistics. Therefore we can use the same argument as F I[3,4] in reducing Eq. (2.2) to arrive at

$$(P\mathbf{z}_1,P\mathbf{z}_2,\cdots P\mathbf{z}_N|e^{-\beta H}|\mathbf{z}_1,\mathbf{z}_2,\cdots,\mathbf{z}_N) = K_\beta\left(\frac{m'}{2\pi\beta\hbar^2}\right)^{3N/2}\exp\left[-\frac{m'}{2\beta\hbar^2}\sum_i(\mathbf{z}_i-P\mathbf{z}_i)^2\right] \times\rho(\mathbf{z}_1,\cdots,\mathbf{z}_N). \quad (2.5)$$

This is further simplified if $\rho(\mathbf{z}_1,\cdots,\mathbf{z}_N)$ is assumed to

* Present address: Armour Research Foundation of Illinois Institute of Technology, Chicago, Illinois.

[1] R. P. Feynman, Phys. Rev. 91, 1291 (1953), hereafter called F I.

[2] R. Kikuchi, Phys. Rev. 96, 563 (1954).

[3] See Eq. (7) of F I.

[4] Professor Feynman in private communication has emphasized the point he raised in footnote 9 of F I and said that, although the effect of a moving atom in permuting other atoms is not important for the Bose-Einstein case where permutations make no difference, this effect must be considered in more detail if we are to apply these ideas to the Fermi-Dirac case. Although it may simply mean a larger effective mass m', it may also be that expression (2.5) is not sufficiently accurate at low temperatures with any m'. In addition, one must be especially careful in Fermi statistics because the plus and minus contributions of even and odd permutations nearly balance out, so that great precision may be needed to keep the balance representative of the truth. Although (2.5) may be correct, he feels that the arguments of F I justify its use only in the E.B. case.

This comment of Professor Feynman warns that one must accept the calculations in the following sections with critical eyes. The author has decided to publish this work, nevertheless, as he believes the expression (2.5) has as much meaning to the problem of He^3 as the Ising model does to that of ferromagnetism.

vanish except for $\mathbf{z}$'s located on a simple cubic lattice,[5] yielding

$$\int (P\mathbf{z}_1, P\mathbf{z}_2, \cdots, P\mathbf{z}_N | e^{-\beta H} | \mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N) d\mathbf{z}_1 \cdots d\mathbf{z}_N = N! Q_0 \exp[-a \sum_i (\mathbf{x}_i - P\mathbf{x}_i)^2], \quad (2.6)$$

where $\mathbf{x}_i$ indicates a lattice point of a simple cubic lattice, a is a constant[6]:

$$a = m'/(2\beta\hbar^2), \quad (2.7)$$

and Q_0 is a factor introduced in Eq. (2.5):

$$Q_0 = K_\beta [m'/(2\pi\beta\hbar^2)]^{3N/2}. \quad (2.8)$$

This is considered a smoothly varying function of temperature,[7] and is independent of permutations pertinent to the present problem.

From Eqs. (2.4) and (2.6), one obtains for the partition function

$$Q_{FD} = Q_0 \sum_P (-1)^P \exp[-a \sum_i (\mathbf{x}_i - P\mathbf{x}_i)^2]. \quad (2.9)$$

The technique used by Kac and Ward[8] in the treatment of the two-dimensional Ising model suggests that Eq. (2.9) can be written in a determinantal form as follows[9]:

$$q \equiv Q_{FD}/Q_0 = \det \mathbf{A}, \quad (2.10)$$

where an element of the matrix $\mathbf{A}$ is defined as

$$\mathbf{A}(\mathbf{x},\mathbf{x}') = \exp[-a(\mathbf{x}-\mathbf{x}')^2], \quad (2.11)$$

$\mathbf{x}$ and $\mathbf{x}'$ taking values of lattice points on a simple cubic lattice. When N is the number of lattice points in the volume or the system, $\mathbf{A}$ is an $N \times N$ matrix. Following Newell and Montroll,[8] one transforms the matrix $\mathbf{A}$ with a unitary matrix $\mathbf{U}$ defined by

$$\mathbf{U}(\mathbf{x},\mathbf{k}) = N^{-\frac{1}{2}} e^{2\pi i \mathbf{x}\cdot\mathbf{k}}, \quad (2.12)$$

obtaining another matrix $\mathbf{B}$ such that

$$\mathbf{B} = \mathbf{U}\mathbf{A}\mathbf{U}^{-1}. \quad (2.13)$$

Then

$$\det \mathbf{A} = \det \mathbf{B}. \quad (2.14)$$

$\mathbf{x}$ and $\mathbf{k}$ in Eq. (2.12), written in components, have the following forms:

$$\mathbf{x} = (k,l,m) \times d, \quad \mathbf{k} = (\kappa,\lambda,\mu) \times (Ld)^{-1}, \quad (2.15)$$

where d is the lattice constant, and $L = N^{\frac{1}{3}}$. k, l, m, κ, λ, and μ in Eq. (2.15) are integers. Because of the lattice structure and the periodic boundary condition with a period of Ld, each of these integers takes one of the values between $-L/2$ and $L/2$.

Combining Eqs. (2.11) and (2.12) and taking account of the periodicity of the lattice, one has

$$\mathbf{B}(\mathbf{k},\mathbf{k}') = \delta(\mathbf{k},\mathbf{k}') \sum_{\mathbf{x}} \mathbf{A}(0,\mathbf{x}) e^{-2\pi i \mathbf{k}\cdot\mathbf{x}}, \quad (2.16)$$

where the summation over $\mathbf{x}$ is to be carried over all the lattice points. Putting Eq. (2.11) into Eq. (2.16) and replacing $\mathbf{x}$ and $\mathbf{k}$ by Eq. (2.15), one obtains

$$\mathbf{B}(\mathbf{k},\mathbf{k}) = \sum_{k,l,m=-L/2}^{L/2} \exp[-ad^2(k^2+l^2+m^2) - 2\pi i(\kappa k + \lambda l + \mu m)/K] = \vartheta_3\left(\frac{\kappa}{K}; e^{-\tau}\right) \vartheta_3\left(\frac{\lambda}{K}; e^{-\tau}\right) \vartheta_3\left(\frac{\mu}{K}; e^{-\tau}\right), \quad (2.17)$$

where ϑ_3 is one of the theta functions and is defined as[10]

$$\vartheta_3(x;t) \equiv \sum_{n=-\infty}^{\infty} t^{n^2} e^{2\pi i n x}, \quad (2.18)$$

and

$$\tau = ad^2 = m'd^2kT/(2\hbar^2). \quad (2.19)$$

In arriving at the ϑ_3-function in Eq. (2.17), L is assumed very large, effectively infinite.

One introduces F by

$$q = e^{-F/kT}, \quad (2.20)$$

uses Eqs. (2.10), (2.14), and (2.17), transforms the summation into an integral, and obtains

$$F/kT = -3N \int_{-\frac{1}{2}}^{\frac{1}{2}} dx \ln \vartheta_3(x; e^{-\tau}). \quad (2.21)^{11,12}$$

This is the part of the free energy which is pertinent to the Fermi-Dirac statistics. As τ is proportional to temperature and $\vartheta_3(x; e^{-\tau})$ is a well-behaved function of its arguments, one can conclude from Eq. (2.21) that this substance does not show any phase change.

It should be noticed that after Eq. (2.9) no approximation is made except the reasonable one that $L \gg 1$. Particularly, it seems worth while to mention here clearly that, although in the previous treatment[2] of He^4 only polygons having sides of length d as shown in Fig. 1(a) was taken into account, in the present paper all of the sides of length longer than d such as shown in Fig. 1(b) are also taken account of.

As the form of Eq. (2.12) suggests, what we have

[5] For the sake of simplicity, the explanation and calculation are based on a simple cubic lattice in the first part of the paper. The face-centered cubic lattice is treated in the later part.

[6] This a is different from a used in reference 2 by a factor of d^2/T.

[7] See p. 1296 of F I.

[8] M. Kac and J. C. Ward, Phys. Rev. **88**, 1332 (1952); G. F. Newell and E. W. Montroll, Revs. Modern Phys. **25**, 353 (1953).

[9] For Bose particles, the partition function q is written as a permanent of $\mathbf{A}$ instead of a determinant as in Eq. (2.10). That is why the technique shown below cannot be applied to He^4.

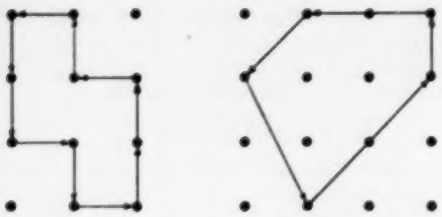
[10] See, for instance, E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, Cambridge and the McMillan Company, New York, 1946), Chap. 21.

[11] D. Lieberman of California Institute of Technology has derived essentially the same formula independently.

[12] Using relations satisfied by the theta functions, Eq. (2.21) can be reduced to a simpler form:

$$F/NkT = \ln 2 - (\tau/4) - \ln[\vartheta_2(0; e^{-\tau})\vartheta_3(0; e^{-\tau})\vartheta_4(0; e^{-\tau})],$$

where ϑ_2, ϑ_3, and ϑ_4 follow the notations of Whittaker and Watson, reference 10.



(a) (b)

FIG. 1. Examples of polygons on a lattice. Each arrow indicates a side of the polygon.

done in this section was to evaluate the partition function using momentum space representation, although we started from the coordinate representation introduced by Feynman. This situation becomes clearer, when one recalls the alternative way of writing the ϑ_3 function:

$$\vartheta_3(x;e^{-\tau})=\left(\frac{\pi}{\tau}\right)^{\frac{1}{2}}\sum_{n=-\infty}^{\infty}\exp[-\pi^2(x+n)^2/\tau]. \quad (2.22)$$

Inserting this into Eq. (2.17), one obtains

$$\mathbf{B}(\mathbf{k},\mathbf{k})=\left(\frac{h^2}{2\pi m'd^2kT}\right)^{\frac{1}{2}}\times\sum_{\mathbf{n}}\exp[-h^2(\mathbf{k}+\mathbf{n}/d)^2/2m'kT], \quad (2.23)$$

where $\mathbf{n}$ is a vector having integers (plus, minus and zero) as its components. Comparing Eq. (2.23) with the momentum space approach of the ideal Fermi particles, one may tell the similarity and also the difference between the two.

At any rate, Eq. (2.21) does not have much physical meaning, because it is for the case of all parallel spins. We are more interested in the effect of plus and minus spins superposed on the Fermi-Dirac statistics and this will be treated in the following sections.

3. FERMI-DIRAC PARTICLES WITH A SPIN 1/2

In order to treat He^3, one has to take into account the fact that an atom possesses nuclear spin of $\frac{1}{2}$. Then the partition function of the previous section has to be revised so as to include spin functions. Extending Eq. (2.4), one sees that the partition function Q is written as

$$Q=N^{-1}\sum_{\text{distr}}\sum_{P_\alpha}\sum_{P_\beta}(-1)^{P_\alpha+P_\beta}\int(P\mathbf{x}_1,P\mathbf{x}_2\cdots P\mathbf{x}_N\times|e^{-\beta H}|\mathbf{x}_1,\cdots,\mathbf{x}_N)d\mathbf{x}_1\cdots d\mathbf{x}_N, \quad (3.1)$$

where P_α (or P_β) is a permutation among atoms having α (or β) spins and $\sum_{\text{distr}}$ is the summation over different distributions of α and β spins over the lattice. P stands for $P_\alpha P_\beta$. Using Eq. (2.5) and assuming as before that the function ρ has nonvanishing values only when its arguments are on a simple cubic lattice, Q is reduced to

$$Q=Q_0\sum_{\text{distr}}\sum_{P_\alpha}\sum_{P_\beta}(-1)^{P_\alpha+P_\beta}\times\exp[-a\sum_j(\mathbf{x}_j-P_\alpha P_\beta\mathbf{x}_j)^2], \quad (3.2)$$

where Q_0 is the same as Eq. (2.8).

Applying the same technique as in the previous section, Eq. (3.2) can be transformed into a determinantal form as

$$q=Q/Q_0=\sum_{\text{distr}}\det\mathbf{A}, \quad (3.3)$$

where an element of the matrix $\mathbf{A}$ is

$$\mathbf{A}(\mathbf{x},\mathbf{x}')=f(\mathbf{x},\mathbf{x}')\exp[-a(\mathbf{x}-\mathbf{x}')^2], \quad (3.4)$$

with $f(\mathbf{x},\mathbf{x}')$ defined by

$$f(\mathbf{x},\mathbf{x}')=\begin{cases}1, & \text{when atoms on } \mathbf{x} \text{ and } \mathbf{x}' \text{ have the same spin,}\\ 0, & \text{when atoms on } \mathbf{x} \text{ and } \mathbf{x}' \text{ have different spins.}\end{cases} \quad (3.5)$$

As the next step, let us work on one of the terms in the summation of Eq. (3.3). Although it is not indicated explicitly, it should be remembered that a term, $\det\mathbf{A}$, corresponds to a certain distribution of spins over the lattice. To evaluate $\det\mathbf{A}$, one transforms the matrix $\mathbf{A}$ by the unitary matrix $\mathbf{U}$ defined in Eq. (2.12), obtaining another matrix $\mathbf{B}$ of Eq. (2.13), whose element is

$$\mathbf{B}(\mathbf{k}',\mathbf{k})=N^{-1}\sum_{\mathbf{x}}\sum_{\mathbf{x}'}e^{2\pi i\mathbf{x}'\cdot\mathbf{k}'}\mathbf{A}(\mathbf{x}',\mathbf{x})e^{-2\pi i\mathbf{x}\cdot\mathbf{k}}. \quad (3.6)$$

Here $\mathbf{x}$ and $\mathbf{x}'$ run through all the lattice points. Now divide the lattice points into groups so that the lattice points in one group have the same distribution of spins in its neighborhood (in its nearest neighbor points or in its nearest and next nearest points, etc., depending on the approximations one chooses). Then

$$\begin{aligned}\mathbf{B}(\mathbf{k}',\mathbf{k})&=N^{-1}\sum_C\sum_{\mathbf{x}'\in C}\sum_{\mathbf{x}}e^{2\pi i\mathbf{x}'\cdot(\mathbf{k}'-\mathbf{k})}\times\mathbf{A}_C(0,\mathbf{x}-\mathbf{x}')e^{-2\pi i(\mathbf{x}-\mathbf{x}')\cdot\mathbf{k}}\\&=N^{-1}\sum_C\Big[\sum_{\mathbf{x}'\in C}e^{2\pi i\mathbf{x}'\cdot(\mathbf{k}'-\mathbf{k})}\Big]\times\sum_{\mathbf{x}''}\mathbf{A}_C(0,\mathbf{x}'')e^{-2\pi i\mathbf{x}''\cdot\mathbf{k}}, \quad (3.7)\end{aligned}$$

where $\mathbf{x}'\in C$ indicates $\mathbf{x}'$ is a member of a group C and $\sum_C$ the summation over different groups. $\mathbf{A}_C(0,\mathbf{x}-\mathbf{x}')$ denotes $\mathbf{A}(0,\mathbf{x}-\mathbf{x}')$ when the origin 0 is a member of the group C. Now we make use of a relation

$$\sum_{\mathbf{x}\in C}e^{2\pi i\mathbf{x}\cdot\mathbf{k}}=Np(C)\delta(\mathbf{k},0), \quad (3.8)$$

where $p(C)$ is the probability of finding a lattice point belonging to the group C. When the points belonging to the group C are scattered at random all over the

lattice, this relation (3.8) is justified as is shown in Appendix. Equations (3.7), (3.8), and (3.4) give

$$\mathbf{B}(\mathbf{k}',\mathbf{k})=\delta(\mathbf{k}',\mathbf{k})\sum_C p(C) \times\sum_{\mathbf{x}}\exp(-ax^2-2\pi i\mathbf{x}\cdot\mathbf{k})f_C(0,\mathbf{x}) =\delta(\mathbf{k}',\mathbf{k})\sum_{\mathbf{x}}\exp(-ax^2-2\pi i\mathbf{x}\cdot\mathbf{k})p(\mathbf{x}), \quad (3.9)$$

where $f_C(0,\mathbf{x})$ means $f(0,\mathbf{x})$ when the origin 0 belongs to a group C, and

$$p(\mathbf{x})=\sum_C p(C)f_C(0,\mathbf{x}) \quad (3.10)$$

is equal to the probability of finding α atoms both at the origin and at the point $\mathbf{x}$, plus the probability of finding β atoms both at the origin and at $\mathbf{x}$.

As is seen in the following sections, a distribution of plus and minus spins over the lattice is specified by a set of parameters. Then, Eq. (3.3) can be written as

$$q=\sum_{\text{distr}}\det\mathbf{B}=\sum_D G[D]\prod_{\mathbf{k}}\mathbf{B}(\mathbf{k},\mathbf{k}), \quad (3.11)$$

where D indicates a set of parameters specifying a distribution and $G[D]$ is a number of different distributions having the same set of values D. In Eq. (3.11), $\mathbf{B}(\mathbf{k},\mathbf{k})$ is a function of D through $p(\mathbf{x})$. If one applies an external field H, the energy of the system increases

TABLE I. Definition of parameters, x_i's, for a lattice point.

Configuration	Probability
α	x_1
β	x_2

by $M[D,H]$ which is a function of D and the field H. When one is interested only in the nuclear magnetic susceptibility, the magnetic energy $M[D,H]$ is due to the nuclear spin magnetic moment, which is assumed to commute with the rest of the Hamiltonian which depends on the coordinates of particles. Therefore the effect of the external field H on q of Eq. (3.11) is multiplication by a factor $\exp(-M[D,H]/kT)$. One evaluates q of Eq. (3.11) by its maximum term, or one maximizes

$$-F/kT=-M[D,H]/kT+\ln G[D] +\sum_{\mathbf{k}}\ln\mathbf{B}(\mathbf{k},\mathbf{k}), \quad (3.12)$$

with respect to the set of parameters D. F in Eq. (3.12) is the part of the free energy we are concerned with.

For the next procedure, it is convenient to rewrite $\mathbf{B}(\mathbf{k},\mathbf{k})$ of Eq. (3.9) using Eq. (2.17) as follows:

$$\mathbf{B}(\mathbf{k},\mathbf{k})=\sum_{\mathbf{x}}\exp(-ax^2-2\pi i\mathbf{x}\cdot\mathbf{k})[p(\mathbf{x})-p_\infty] +p_\infty\vartheta_3\left(\frac{\kappa}{K};e^{-\tau}\right)\vartheta_3\left(\frac{\lambda}{K};e^{-\tau}\right)\vartheta_3\left(\frac{\mu}{K};e^{-\tau}\right). \quad (3.13)$$

where

$$p_\infty=\lim_{|\mathbf{x}|\to\infty}p(\mathbf{x}). \quad (3.14)$$

τ in Eq. (3.13) is defined in Eq. (2.19) and is proportional to temperature.

TABLE II. Definition of parameters, y_i's, for a bond. ν_i is the number of configurations having the same probability y_i.

Configuration	Probability	ν_i
$\alpha-\alpha$	y_1	1
$\alpha-\beta$	y_2	2
$\beta-\beta$	y_3	1

4. SIMPLE CUBIC LATTICE TREATMENT

In order to proceed further, one chooses a set of parameters to specify distributions. The simplest approximate treatment is explained in this section. The procedure follows closely that derived by the author in the treatment of the Ising model,[13] and readers are referred to the original paper if necessary.

In the first approximation, to be called the "pair approximation," the parameters one uses are probabilities of finding configurations on lattice points and those on bonds as shown in Tables I and II. From geometrical consideration, one derives the relations among the parameters as shown in Table III.

For the weight factor $G[D]$ of Eq. (3.12), one uses the last equation of Sec. B of TCP I, or

$$G=X_N{}^2Y_N{}^{-2}N!^{-2}, \quad (4.1)[14]$$

where

$$X_N=\prod_{i=1}^{2}(x_iN)!, \quad Y_N=\prod_{i=1}^{3}(y_iN)!^{\nu_i}. \quad (4.2)$$

Hence, using Stirling's formula,

$$\ln G=N\left[5\sum_{i=1}^{2}x_i\ln x_i-3\sum_{i=1}^{3}\nu_iy_i\ln y_i\right]. \quad (4.3)$$

Using the definition of $p(\mathbf{x})$ and Table II, one derives

$$\begin{aligned}&p(\text{origin})=1,\\ &p(\text{nearest neighbor})=y_1+y_2,\\ &p(\text{further neighbor})\doteqdot p_\infty=x_1{}^2+x_2{}^2.\end{aligned} \quad (4.4)$$

The last of these relations is approximate, but seems consistent with the approximation one makes for G in

TABLE III. Relations among parameters for the simple cubic lattice treatment. The meaning of the table is, for instance, $y_1=\frac{1}{3}+\xi_1-y_3$. ξ_1 and y_3 are independent variables.

	1	ξ_1	y_3
x_1	$\frac{1}{3}$	1	
x_2	$\frac{1}{3}$	−1	
y_1	$\frac{1}{3}$	1	−1
y_2	$\frac{1}{3}$	−1	−1

[13] R. Kikuchi, Phys. Rev. 81, 988 (1951), hereafter called TCP I.

[14] This formula is also discussed in Kurata, Kikuchi, and Watari, J. Chem. Phys. 21, 434 (1953). Equation (4.1) gives Bethe's approximation when applied to the Ising model.

Eq. (4.1). The magnetic energy of the system is equal to

$$M[D,H]=\mu HN(x_2-x_1)=-2\mu HN\xi_1, \quad (4.5)$$

where μ is the nuclear magnetic moment.

Inserting Eqs. (4.3), (4.4), and (4.5) into Eqs. (3.12) and (3.13), and transforming the summation into an integral, one obtains

$$-\frac{F}{NkT}=\frac{2\xi_1\mu H}{kT}+\frac{\ln G}{N}+\int d\mathbf{r}\ln\Omega, \quad (4.6)$$

where

$$\Omega=(\tfrac{1}{2}+2\xi_1^2)\vartheta_3(x;\epsilon^{-\tau})\vartheta_3(y;\epsilon^{-\tau})\vartheta_3(z;\epsilon^{-\tau})+\tfrac{1}{2}$$
$$-2\xi_1^2+(\tfrac{1}{2}-2y_2-2\xi_1^2)2\epsilon^{-\tau}\sum\cos 2\pi x, \quad (4.7)^{15}$$

with

$$\sum\cos 2\pi x\equiv\cos 2\pi x+\cos 2\pi y+\cos 2\pi z, \quad (4.8)$$

and

$$\int d\mathbf{r}\equiv\int_{-\frac{1}{2}}^{\frac{1}{2}}dx\int_{-\frac{1}{2}}^{\frac{1}{2}}dy\int_{-\frac{1}{2}}^{\frac{1}{2}}dz. \quad (4.9)$$

Differentiation of Eq. (4.6), with respect to ξ_1, gives

$$5\ln\frac{x_1}{x_2}-3\ln\frac{y_1}{y_2}+2\frac{\mu H}{kT}+\int d\mathbf{r}4\xi_1[\vartheta_3(x)\vartheta_3(y)\vartheta_3(z)$$
$$-1-2\epsilon^{-\tau}\sum\cos 2\pi x]/\Omega=0. \quad (4.10)^{16}$$

When H is zero, $\xi_1=0$ is a solution of this equation.[15] Another possibility, $\xi_1\neq 0$, corresponds to a ferromagnetic state, to which we shall refer later, but first we solve the problem of the nonferromagnetic case. As we are interested in the magnetic susceptibility which is the property defined for vanishing magnetic field, the second equation $\partial F/\partial y_2=0$ is to be solved for y_2 under the condition $\xi_1=0$. Then one obtains

$$3\ln\frac{1-2y_2}{2y_2}=\int_{\Omega_0}\frac{d\mathbf{r}}{-2\epsilon^{-\tau}\sum\cos 2\pi x}, \quad (4.11)$$

where Ω_0 is Ω of Eq. (4.7) with $\xi_1=0$. The magnetic susceptibility χ is defined from Eq. (4.10) as

$$\frac{1}{\chi}=\lim_{H\to 0}\frac{H}{2\mu\xi_1}=\frac{kT}{\mu^2}\left[\frac{3}{1-2y_2}-5-\int d\mathbf{r}\{\vartheta_3(x)\vartheta_3(y)\vartheta_3(z)\right.$$
$$\left.-1-2\epsilon^{-\tau}\sum\cos 2\pi x\}/\Omega_0\right]. \quad (4.12)^{16}$$

As it seems hopeless to solve Eq. (4.11) analytically, it was solved numerically,[17] to obtain y_2 and hence χ of Eq. (4.12) as functions of τ. Then in order to compare the results with observations, τ was converted to actual temperature. Assuming that the density[18] of He^3 is 0.08 g/cc and is independent of temperature, Eq. (2.19) is written as

$$T=2.0(m/m')\tau, \quad (4.13)$$

where m is the true mass of a He^3 atom, and m' is its effective mass introduced by Feynman in F I. As T is proportional to τ, if one first plots χ against τ, it is easily converted to a function of T, only by changing the temperature scale, the scale factor depending on m'/m. In Fig. 2, $\chi T/C$ and $y_2-0.25$ so obtained assuming $m'/m=5.5$ are plotted against T with solid curves. This value of the ratio is taken in order to give the best fit of the observations by Fairbank, Ard, and Walters[19] which are also shown in Fig. 2 with black circles. Fairly good agreement with the experiments is also obtained for the values of m'/m in the range

$$5.0<(m'/m)_{as}<5.5. \quad (4.14)$$

Comparing the treatment of the Ising model in TCP I and the present calculation, one knows that the ferromagnetic Curie temperature, if it exists, is to be given by the condition: $\chi\to\infty$. From the curve in Fig. 2, it is clear that this condition is not satisfied, and hence there is no ferromagnetic state, within the temperature range of the present calculation.[20]

The value of m'/m in Eq. (4.14) seems to be much too large, when one recalls that the value of the corresponding ratio for He^4 was estimated by Feynman[1] as

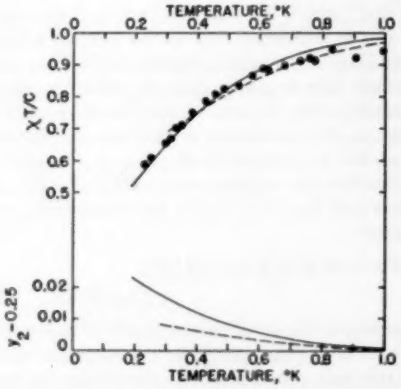


FIG. 2. Plots of nuclear magnetic susceptibility χ and the parameter y_2 defined in Table II against temperature. C in $\chi T/C$ is the normalization constant such that $\lim_{T\to\infty}\chi T/C=1$. The black circles are experiments by Fairbank, Ard, and Walters.[19] The solid curves are the calculated ones based on the simple cubic lattice treatment ("pair" and "square" approximations) and $m'/m=5.5$. The broken curves are the calculated ones based on the face-centered cubic lattice treatment ("tetrahedron" approximation) and $m'/m=3.5$.

[15] ξ_1, introduced in Table III, is a long-range order parameter. x, y, and z without subscripts are cartesian coordinates and are to be distinguished from the parameters, x_i's and y_i's, defined in Tables I and II.

[16] $\vartheta_3(x;\epsilon^{-\tau})$ is simply written as $\vartheta_3(x)$, dropping the second argument $\epsilon^{-\tau}$, as it would not cause any confusion.

[17] For the numerical integration, Weddle's rule was used. This is to approximate $\int_{-3}^{3}f(x)dx$ by $(3/10)[f(-3)+5f(-2)+f(-1)+6f(0)+f(1)+5f(2)+f(3)]$. See, for instance, D. R. Hartree, *Numerical Analysis* (Oxford University Press, Oxford, 1952), p. 101.

[18] E. C. Kerr, Phys. Rev. **96**, 551 (1954).

[19] Fairbank, Ard, and Walters, Phys. Rev. **95**, 566 (1954).

[20] L. Goldstein and M. Goldstein, J. Chem. Phys. **18**, 538 (1950).

about 1.5 and was 1.3 by the author's calculation.[3] It was suspected that this ratio might change when one improves the approximation and uses a square as the basic figure, just as was done for the Ising model in Sec. D of TCP I. But the actual calculation based on the "square" approximation gave practically the same curve[21] for $\chi T/C$ as a function of τ. This shows on one hand that the result shown in Fig. 2 is practically the rigorous solution of the model we used in this section, and on the other that the value of m'/m in (4.14) is inherent in the model itself and does not depend on the approximation used in this section.

Therefore, as the next step to examine m'/m, the face-centered cubic lattice is taken with a hope that it might give a lower value of this ratio. It will be explained in the next section.

5. FACE-CENTERED CUBIC LATTICE TREATMENT

The calculation based on the face-centered cubic lattice is essentially the same as that for the simple cubic lattice. But in order to make clear the approximation used in the process of the calculation, some of the equations will be listed below.

Equations (2.1) to (2.14) remain valid for this case, but as the reciprocal of a face-centered cubic lattice is a body-centered cubic lattice, Eq. (2.15) should be modified accordingly. Equation (2.17) is replaced by

$$\mathbf{B}(\mathbf{k},\mathbf{k})=\Theta\left(\frac{\kappa}{K},\frac{\lambda}{K},\frac{\mu}{K};e^{-2\tau}\right), \tag{5.1}$$

where

$$\begin{aligned}\Theta(x,y,z;t)\equiv&\vartheta_2(2x;t)\vartheta_2(2y;t)\vartheta_2(2z;t)\\&+\vartheta_2(2x;t)\vartheta_2(2y;t)\vartheta_2(2z;t)\\&+\vartheta_2(2x;t)\vartheta_2(2y;t)\vartheta_2(2z;t)\\&+\vartheta_2(2x;t)\vartheta_2(2y;t)\vartheta_2(2z;t),\end{aligned} \tag{5.2}$$

and $\vartheta_2(x;t)$ is another theta function defined by

$$\vartheta_2(x;t)\equiv\sum_{n=-\infty}^{\infty}t^{(n+\frac{1}{2})^2}e^{\pi i(2n+1)x}. \tag{5.3}$$

Consequently, Eq. (3.13) is changed into

$$\begin{aligned}\mathbf{B}(\mathbf{k},\mathbf{k})=&\sum_{\mathbf{x}}\exp(-a\mathbf{x}^2-2\pi i\mathbf{x}\cdot\mathbf{k})[\rho(\mathbf{x})-\rho_\infty]\\&+\rho_\infty\Theta\left(\frac{\kappa}{K},\frac{\lambda}{K},\frac{\mu}{K};e^{-2\tau}\right).\end{aligned} \tag{5.4}$$

As it is known that the "pair" approximation for the face-centered cubic lattice is poor[22] because the correlation among the first shell points is neglected, the tetrahedron approximation[23] will be taken up in this section. Then in order to proceed further, one uses besides Tables I and II probability parameters defined in Table IV. These parameters, z_i's, are expressed as functions of ξ_1 and y_2 of Table III and additional two variables, ξ_3 and z_3, as shown in Table V.

For the free energy F, the expression (4.6) is used with different definitions for G and Ω:

$$\frac{\ln G}{N}=6\sum_{i=1}^{3}\nu_i y_i\ln y_i-2\sum_{i=1}^{4}\gamma_i z_i\ln z_i-5\sum_{i=1}^{2}x_i\ln x_i, \tag{5.5}$$

and

$$\begin{aligned}\Omega=&(\tfrac{1}{2}+2\xi_1^3)\Theta(x,y,z;e^{-2\tau})+\tfrac{1}{2}-2\xi_1^3\\&+(\tfrac{1}{2}-2y_2-2\xi_1^3)2e^{-\tau}\sum\cos2\pi(x+y),\end{aligned} \tag{5.6}$$

where

$$\begin{aligned}\sum\cos2\pi(x+y)\equiv&\cos2\pi(x+y)+\cos2\pi(y+z)\\&+\cos2\pi(z+x).\end{aligned} \tag{5.7}$$

TABLE IV. Definition of parameters, z_i's, for a tetrahedron. γ_i is the number of configurations having the same probability z_i. This table is the same as Table VII(a) of TCP I with different notations.

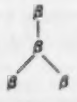
Configuration	Probability	γ_i
	z_1	1
	z_2	4
	z_3	6
	z_4	4
	z_5	1

TABLE V. Relations among parameters. The meaning of the table is the same as Table III.

	1	ξ_1	y_2	ξ_3	z_3
z_1	$\frac{1}{8}$	1	−2	−1	1
z_2			$\frac{1}{2}$	$\frac{1}{2}$	−1
z_4			$\frac{1}{2}$	$-\frac{1}{2}$	−1
z_5	$\frac{1}{8}$	−1	−2	1	1

[3] The deviation was only 0.3% even at the lowest temperature calculated, $\tau=0.5$.

[22] See, for instance, Fig. 9 of TCP I for the two-dimensional triangular net.

[23] This corresponds to Sec. H of TCP I.

Equation (5.5) follows from Eq. (H1.2) of TCP I. The next procedure is to minimize F with respect to independent variables to obtain

$$3\ln\frac{y_2}{y_1}-2\ln\frac{z_2}{z_1}=4e^{-\tau}\int\frac{d\mathbf{r}}{\Omega_0}\sum\cos 2\pi(x+y), \qquad z_1z_2^3=z_2^4, \tag{5.8}$$

corresponding to Eq. (4.11). In place of Eq. (4.12), one has

$$\frac{1}{\chi}=\frac{kT}{\mu^2}\left[5+\frac{14y_2-4}{(1-2y_2)(1-3y_2)}-4\int d\mathbf{r}\{\Theta(x,y,z;e^{-2\tau})-1-2e^{-\tau}\sum\cos 2\pi(x+y)\}/\Omega_0\right]. \tag{5.9}$$

The two equations (5.8) are solved for y_2 and z_2 numerically and these values are inserted in (5.9) to evaluate χ. The results of the calculation are shown in Fig. 2 with dotted curves. In the process of arriving at these curves, the temperature scale was changed from τ to T by the relation

$$T=1.6(m/m')\tau, \tag{5.10}$$

and the value $m'/m=3.5$ was used. The range of the value of this ratio which gives fairly good fit with the experiments is

$$3.2<(m'/m)_{tet}<3.8. \tag{5.11}$$

In order to make sure of these numbers, a calculation taking into account the second and the third neighbors was carried out. It was found that this improved approximation gave essentially the same values as in (5.11), and one can conclude, for this lattice also, that the values in (5.11) are very close to the rigorous results of the model.

6. DISCUSSION

Equations (4.14) and (5.11) show that $(m'/m)_{tet}$ is smaller than $(m'/m)_{sc}$. This can be interpreted as follows. As one sees from the lower curves of Fig. 2, y_2 is greater than $\frac{1}{4}$. This means that $\alpha-\beta$ (and $\beta-\alpha$) pairs appear more often than $\alpha-\alpha$ and $\beta-\beta$ pairs; in other words the system shows an "antiferromagnetic" tendency. Wannier[35] proved that the Ising model of a two-dimensional triangular net does not exhibit antiferromagnetic state whereas the square net[36] does. This is interpreted as due to the fact that triangular net is divided into three equivalent sublattices whereas the square net divides into two. The same argument holds for the three-dimensional case and one expects that the antiferromagnetic state is difficult to be realized for the face-centered cubic lattice which is divided into four sublattices whereas it is easier for the simple cubic lattice which has two sublattices. This is the reason why one obtains smaller values of y_2 and hence smaller deviation of $\chi T/C$ from unity for the face-centered cubic lattice than the simple cubic, corresponding to the same value of τ. The larger the $\chi T/C$ is, the lower the m'/m becomes.

Another factor to lower $(m'/m)_{tet}$ compared to $(m'/m)_{sc}$ is the change of the numerical factors, 2.0 and 1.6, of Eqs. (4.13) and (5.10), respectively, which are inversely proportional to square of the hypothetical lattice constant, as is seen in Eq. (2.19). When one assumes no holes, the lattice constant for the face-centered cubic lattice is larger than that for the simple cubic in order to give the same macroscopic density. The smaller this numerical factor is, the smaller the m'/m becomes.

Even though one knows the values of (4.14) and (5.11) for the two regular lattices, for the present there is no practical way of calculating the value of m'/m for a random distribution of atoms, though probably it is not too unreasonable to guess that the true value would lie in between the two.

The final problem which still remains is why the value of m'/m should be so large (3.2∼5.5) in order to fit the observation. Probably the answer lies in the passage from Eq. (2.4) to Eq. (2.5) as Feynman pointed out,[4] but it will not be discussed further in this paper.

It is certainly of much importance to calculate the entropy of the system in order to discuss the nature of the approximation and also to check Pomeranchuk's prediction,[37] but the entropy which we can calculate using the method of this paper is only related to q of Eq. (3.3) and is not the total entropy of the system, as the factor Q_0 of that equation is unknown. Since the interpretation of the entropy derived from q is not settled yet, it will not be reported here.

As a conclusion we can say that the partition function which Feynman proposed can explain at least qualitatively the properties of liquid He^4 and also of liquid He^3 without contradiction.

ACKNOWLEDGMENTS

Thanks are due Professor R. P. Feynman, Professor J. E. Mayer, and Professor M. H. Cohen for enlightening discussions.

[34] These are the equations obtained by minimizing the free energy F with respect to y_2 and z_2 and then putting ξ_1 and ξ_2 equal to zero. The other two equations, which correspond to Eq. (4.10) and are derived by minimizing F with respect to ξ_1 and ξ_2 were not listed here, though Eq. (5.9) is derived from these two equations. As was mentioned in connection to Eq. (4.10) we know that $\xi_1=0$ and $\xi_2=0$ are possible solutions of these unlisted two equations for the vanishing value of H, and we are interested only in these values of ξ's. ξ_1 and ξ_2 are both long-range order parameters and their nonvanishing values correspond to ferromagnetic state which does not occur in our problem.

[35] G. H. Wannier, Phys. Rev. **79**, 357 (1950).

[36] L. Onsager, Phys. Rev. **65**, 117 (1944).

[37] I. Pomeranchuk, J. Exptl. Theoret. Phys. (U.S.S.R.) **20**, 1919 (1950); Osborne, Abraham, and Weinstock, Phys. Rev. **94**, 202 (1954); L. Goldstein, Phys. Rev. **96**, 1455 (1954).

APPENDIX. DISCUSSION OF EQ. (3.8)

For simplicity let us look at the part of Eq. (3.8) related to the x-component. Then the problem is to evaluate the value of the function

$$g(\kappa)=\sum_{k\in C} e^{2\pi i k\kappa/K}, \qquad (A.1)$$

where k and κ are the same as in Eq. (2.15). In the complex number plane, $g(\kappa)$ is the vector connecting the origin to the end point of the succession of vectors of the summand. When one assumes that points belonging to the group C are scattered at random, the problem is looked upon as random walk in the complex number plane, and the probability $\bar{W}(\mathbf{r})d\mathbf{r}$ of finding the vector $g(\kappa)$ lying within the interval $(\mathbf{r}, \mathbf{r}+d\mathbf{r})$ from the origin is given by

$$\bar{W}(\mathbf{r})d\mathbf{r}=(\pi n)^{-1}\exp(-|\mathbf{r}|^2/n)d\mathbf{r}. \qquad (A.2)^{28}$$

In this equation n is the total number of arrows considered, i.e., $n=N\bar{p}(C)$. As $\bar{p}(C)$ is independent of N, when N becomes larger $\bar{W}(\mathbf{r})$ approaches to the delta function, $\delta(\mathbf{r})$, closer and closer. Therefore for very large N, Eq. (3.8) is justified.

[28] See, for instance, S. Chandrasekhar, Revs. Modern Phys. 15, 1 (1943).

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Photoelectric Mixing of Incoherent Light*

A. Theodore Forrester,† Richard A. Gudmundsen,‡ and Philip O. Johnson§
University of Southern California, Los Angeles, California

(Received April 20, 1955)

Beats have been obtained between incoherent light sources by mixing Zeeman components of a visible spectral line at a photosurface. Periodicity in emission was observed through the excitation of a 3-cm cavity. Because of incoherence between the spectral lines and incoherence between the beats from different photocathode areas, the signal-to-shot-noise ratio at the cavity is only 3×10^{-8} but the beats were modulated optically, while maintaining constant total intensity and our receiver was able to yield a signal-to-noise ratio of two at the indicator. The basic idea is that, in the photoelectric process, the emission probability for electrons is proportional to the square of the resultant electric field amplitude, implying an interference between light originating in independent sources. This is a point of view which does not appear to be tested in any other experiment involving quantum effects. The experiment also demonstrates that any time delay between photon absorption and electron release must be significantly less than 10^{-10} second.

I. INTRODUCTION

THE combination of two wave trains of slightly different frequencies is equivalent to a wave of the average frequency modulated by the difference frequency. This is evidenced in the phenomenon of acoustical beats and is responsible for the operation of superheterodyne radio receivers. The periodic variation in intensity which occurs at a fixed point on the image of a Michelson interferometer when one of the mirrors is moving may also be interpreted as beats between the light reflected from the stationary mirror and light which has had its frequency changed by reflection from the moving mirror.[1–3] However, the problem of beating incoherent light waves, i.e., light waves which originate in different sources, is quite different, and since the publication of the original suggestion,[4,5] it has been argued[6] that the observation of such beats is impossible. These arguments, when examined carefully, really provide reasons why beats between incoherent light sources are difficult, rather than impossible, to detect. Were optical lines much sharper than they are, or possible to produce, without great broadening, in much greater intensity than present techniques permit, beats between incoherent light waves would be easy to observe.

Following the publication of the original suggestion[4] for this experiment, Ruark,[7] calling attention to an

* Supported in part by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

† Now on leave at Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania.

‡ Now at Hughes Aircraft Corporation, Culver City, California.

§ Now at North American Aviation, Downey, California.

[1] A. Righi, J. Physique 2, 437 (1883) describes an ingenious production of light beats. To demonstrate that light which has passed through a rotating Nichol prism may be resolved into two circularly polarized beams, one increased and one decreased in frequency with respect to the incident light, he performed an experiment, in its essence a double-slit interference experiment, in which one slit was illuminated by light of reduced frequency and the other slit by the increased frequency, both altered to be plane polarized in the same plane. The moving fringe pattern he observed, and interpreted as beats, is, in principle, no different from those produced by moving a mirror of a Michelson interferometer.

[2] E. Rüchardt, Optik 6, 238 (1950), raises an objection to this point of view based on a requirement for coherence over a beat period. His objection is adequately answered in reference 3.

[3] C. V. Fragstein, Optik 8, 289 (1951).

[4] Forrester, Parkins, and Gerjuoy, Phys. Rev. 72, 728 (1947).

[5] Gerjuoy, Forrester, and Parkins, Phys. Rev. 73, 922 (1948).

[6] L. R. Griffith, Phys. Rev. 73, 922 (1948).

[7] A. Ruark, Phys. Rev. 73, 181 (1948).

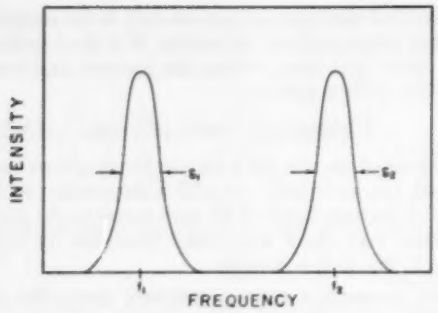


FIG. 1. Spectrum required for the production of sharp beats.

earlier paper,[8] pointed out that waves of different frequencies are not necessarily incoherent, a point of view with which we are in complete agreement. When the mirror of a Michelson interferometer is moved, or in Righi's experiment,[1] the two frequencies are thoroughly coherent. When a periodic modulation of a wave occurs, the sidebands generated in the process have phases which are related and are therefore coherent. We wish to make a special point of the fact that the lines we are mixing are regarded as completely incoherent with respect to each other; i.e., that the relative phase shift of the two lines is completely random and occurs at a rate limited only by the widths of the two lines.

Imagine a source of light emitting two spectral lines arising from two sets of atoms and therefore incoherent. The line widths δ_1 and δ_2, shown in Fig. 1, can arise in many ways, e.g., as a natural width or from Doppler broadening. Whatever the source of the broadening, this curve may be thought of as a plot of the squares of the amplitudes of the Fourier components of the electromagnetic field which is the light. The combination of these two lines should show beats between all of the Fourier components of one line and all of the components of the other (and also between components of a single line which can be ignored because they occur in a different frequency range). The variation in amplitude with time is shown in Fig. 2(a) where the effect of a line width which is not negligible is displayed by irregularities in the wave envelope. The beat frequencies will vary approximately from $f_2-f_1-\delta$ to $f_2-f_1+\delta$ but will nevertheless be a well-defined beat pattern providing that

$$f_2-f_1 \gg \delta, \tag{1}$$

which is equivalent to stating that the coherence time $1/\delta$ should be long compared to the beat period $1/(f_2-f_1)$.

Line widths, in the visible region, for allowed transitions, run about 10^8 cps but because it is hard to avoid a great deal of broadening when high intensity is sought, 10^9 cps is a more reasonable figure to use in making estimates. The inequality (1) then requires that f_2-f_1 be of the order of 10^{10} cps or greater to produce a sharp beat frequency. Fortunately 10^{10} cps is a very convenient region of the electromagnetic spectrum in which to detect energy and is a frequency separation easily produced by the Zeeman effect.

A nonlinear device is required for the generation of the difference frequency. For light waves a solution is provided by the photoelectric effect, in which, over a small range of frequencies, the current is proportional to light intensity, or electric field amplitude squared. Indeed, the assumption that the probability of emission of an electron is proportional to the square of the total electric field intensity is really the fundamental assumption of this experiment. It implies an interference between light originating in different atoms, which many physicists find contradictory to their ideas about the nature of interference.

The validity of this approach to the photoelectric effect, rather than, for example, a squaring of the amplitudes of the Fourier components before addition (which would yield a time-independent current), although extremely basic, does not appear to be tested, at least with anything like this degree of directness, in any other experiment. Furthermore, it seems that in no other experiment involving a quantum effect as detector, is an effect observed which must be interpreted as interference between independently generated waves, including wave functions for, say, electrons.

It is worth noting that the generation of the beat frequency depends on what would ordinarily be called a one-quantum process. Any effect which required that photons act in pairs, or successively, to produce emission is so small compared even to the observed effect, as to be considered, at present, indiscernible.

Makinson[9] has made a calculation which is a more rigorous treatment of the photoelectric effect than we attempted. Since he describes the electric fields classically and permits the addition of waves of different frequencies, it is a calculation which we would expect

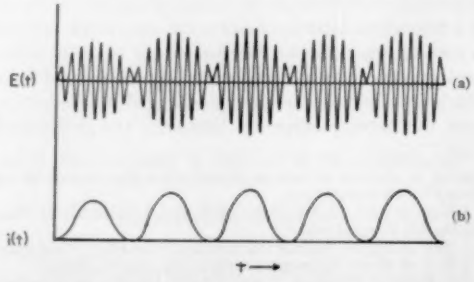


FIG. 2. Electric field variation (a) due to an admixture of two spectral lines and the accompanying photoelectric current (b). The irregularities show the effect of line widths which are not negligibly small.

[8] Breit, Ruark, and Brickwedde, Phil. Mag. 3, 1306 (1927).

[9] R. E. B. Makinson, Phys. Rev. 75, 1908 (1949).

to reduce to our own. However, in his application of his results to the problem of light beats he finds the phase of the beats to vary with the velocity of the emitted electrons which, he states, would result in a very small percentage modulation, whereas the starting point in our calculations is that the photocurrent from a sufficiently small area is 100 percent modulated, as shown in Fig. 2(b). Makinson's conclusion should be based on a demonstration that the variation in phase with velocity is of the order of π or greater. As far as we are aware, the magnitude of this phase variation has not been calculated.

II. ORDER OF MAGNITUDE CALCULATIONS

A. Comparison between Signal and Shot Noise

It is shot noise which provides the basic obstacle to the observation of the process as outlined and a very simple picture suffices for an estimation of the signal-to-shot noise ratio which can be expected. In Fig. 2(b) the photocurrent, resulting from a spectrum made up of two lines satisfying the criterion (1), is shown. Shot noise which would cause variations in current not mirroring those in the field intensity is neglected for the moment. Unfortunately, beats from widely separated areas of the cathode cannot be expected to be in phase, unless we use perfectly plane light. In a more realistic case we can expect the relative phase of each of the light waves and, therefore, the phase of the beats to be constant over an area of the order of the size of the diffraction pattern of a single point on the source,[10] i.e., λ^2/Ω, where Ω is the solid angular spread in the light at the photocathode. From this area the average peak value of the ac current, at the beat frequency, is equal to the average or dc current, so that the mean square value of the ac current from this elementary area is

$$\langle i_0^2\rangle = \tfrac{1}{2} i_{dc}^2 = \tfrac{1}{2}(I\lambda^2/A\Omega)^2, \tag{2}$$

where I is the total photocurrent from a cathode of total area A. If we consider the signal from each area of coherence (meaning an area of the size λ^2/Ω) to be randomly phased with respect to all others, then we must increase the ac current for the entire cathode by a factor $\sqrt{n}$, where $n=A\Omega/\lambda^2$ is the number of times the area of coherence goes into the entire cathode area. Thus for the entire cathode, the mean square current is

$$\langle i^2\rangle = \langle i_0^2\rangle A\Omega/\lambda^2 = I^2\lambda^2/2A\Omega. \tag{3}$$

The dc current is proportional to n, so that, if n is very large, the current has the character of dc current modulated by the minute amount given by Eq. (3). Inevitably associated with a dc photocurrent I is the mean square shot noise current $\langle i_n^2\rangle = 2eI\Delta f$. To compare this with the signal, Δf should be taken as the frequency interval throughout which the signal is spread, approximately the line width δ times $\sqrt{2}$, leading to

$$\langle i^2\rangle/\langle i_n^2\rangle = \lambda^2 I/4\sqrt{2}e\delta A\Omega, \tag{4}$$

an equation which we can use in determining the order of magnitude of the expected signal-to-noise in the emitted current. The experimental signal-to-noise ratio, aside from errors arising from the character of the approximations made in deriving Eq. (4), will be worse than this due to receiver noise. However when circuit noise is minimized and I made as large as inequality (1) allows, the shot noise does become the major source of noise.

Using the green line $\lambda = 5.461\times10^{-5}$ cm of Hg^{198}, and, after passing the light through an optical system containing a filter to remove all Hg lines but this one and a polaroid to select the appropriate components of the Zeeman spectra, we were able to get a photocurrent of 3.85×10^{-6} ampere from an $A\Omega$ of 0.17 cm^2-steradian, under conditions for which $\delta = 8\times10^8$ cps.[11] These numbers in Eq. (4) lead to an expected signal-to-noise ratio of approximately 10^{-4}.

B. Observability of this Effect

The detection of an effect so heavily overwhelmed by noise demands that a modulation be imposed on the signal in such a way that the noise remains unmodulated. In a technique which is now common[12] the modulated signal, after detection, is passed through a very narrow band amplifier, preferably phase selective, before registering on the indicator. A system of this kind gives a signal-to-noise ratio at the final indicator which is,[13] approximately,

$$S/N = \epsilon(\Delta_1 f/\Delta_2 f)^{\frac{1}{2}}, \tag{5}$$

where ϵ is the input signal-to-noise ratio, $\Delta_1 f$ the band width of the input, and $\Delta_2 f$ the band width of the narrow-band amplifier.

Unfortunately, it is not feasible to choose $\Delta_1 f$ equal to the spectral line width, approximately 1000 megacycles per second. It was found difficult to make a very low-noise microwave receiver with a band width greater than 7 megacycles. Since we arrived at $\epsilon = 10^{-4}$ in Sec. IIA, we require here, for $S/N = 1$ that $\Delta_2 f = 0.07$ cps. Actually effects not taken into account in this rough calculation, such as the effect of extraneous light on the photocathode, worsen the situation so that, in practice, a response time of 250 seconds, implying $\Delta_2 f \approx 0.004$, was able to give a signal-to-noise ratio of only two at the final indicator. A great deal of patience is required to obtain data under these conditions and pains have to be taken to see that effects of amplifier drift do not give spurious signals.

[10] Discussed at greater length in Sec. IVB.

[11] Factors governing the choice of this spectral line, and conditions under which these numbers were obtained are discussed in Sec. IIID.

[12] R. H. Dicke, Rev. Sci. Instr. 17, 268 (1946).

[13] Easily seen to be an approximation to the exact Eq. (17), as shown following Eq. (22). A qualitative argument is presented by R. H. Dicke (reference 12).

The accurate quantitative calculation from which Eqs. (4) and (5) may be found as approximations, is deferred to Sec. IV so that it can be developed with specific reference to the circuitry presented in Sec. IIID.

III. EXPERIMENTAL EQUIPMENT AND PROCEDURES

A. Modulation Technique

While producing the modulation of the signal it is necessary that the shot noise remain unmodulated to the order of one part in 10^5. It is not possible, for example, to produce the desired modulation by modulating the magnetic field responsible for the line separation because the light intensity is too dependent on the magnetic field intensity. It did prove possible to modulate with constant total intensity by taking advantage of the difference in the polarization of the central and outer components of the Zeeman spectrum with the optics shown in the upper left of Fig. 3. Ignoring, for the moment, the optics to the left of the light source, light from the source passes through a rotating half-wave plate whose effect on a polarized beam of light is to cause a rotation of the plane of polarization at twice the rate at which the retardation plate rotates. The action on polarized light of the rotating $\lambda/2$ plate followed by a polaroid is to produce an interruption of the light at four times the rotation frequency. Viewed perpendicular to the magnetic field, the light is composed of the π components polarized parallel to the magnetic field, and the σ components, polarized normal to the magnetic field so that the effect of the rotating $\lambda/2$ plate plus polaroid B is to produce an alternation between the π and σ components at the photosurface. If the magnetic field is set so that beats between σ components occur at the frequency of the cavity, then this will result in a modulation of these beats.

If the total light from the source is unpolarized this modulation will occur with no variation in total light intensity or photoelectric emission. However, it was found that our sources were never unpolarized to a degree even approaching one part in 10^3. Even in the absence of a magnetic field a light source is usually polarized by an amount of the order of 1 percent. In a magnetic field the polarization of our microwave excited electrodeless discharges was of the order of 10 percent,[14] or 10^4 times too large. An adjustable tipped glass plate which could compensate for 6–8 percent polarization was placed between the source and the rotating $\lambda/2$ plate as shown in Fig. 3. The remainder of the polarization compensation was accomplished by making use of the light leaving the source in the reverse direction. This light was polarized, adjusted in intensity and plane of polarization and sent back through the source so as to produce zero signal on the compensation monitoring oscilloscope and zero signal on the recording meter which is the output of the compensation monitoring circuit chain.

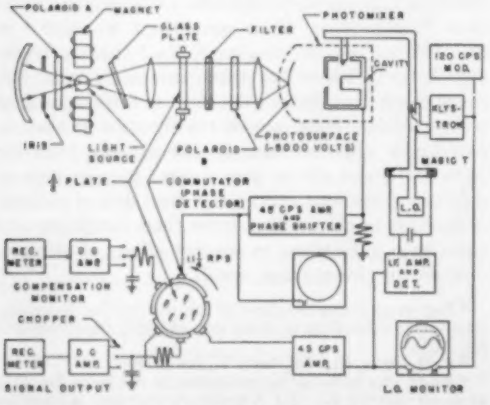


FIG. 3. Basic apparatus for the photoelectric mixing experiment.

B. Photomixer

It is in the photoelectric mixing tube, or photomixer, shown in Fig. 4, that the beat frequency is generated. Because of the small signal, even under the most favorable conditions, it is important that the photosurface be as sensitive as possible. The other requirements for the tube can be quantitatively obtained from the calculations of Sec. IVA, but it is easy to understand qualitatively what they are. The cathode must have the largest possible area and see the largest solid angle so that the photocurrent can be large. The signal-to-shot-noise ratio is not affected by this increase in total current but it is important in order to get above detector noise. For the same reason it is important that the resonant cavity in which the current pulses are converted to electrical energy receive almost all of the current and have as high a Q and R (shunt resistance) as possible. The latter requirement calls for a large gap in the cavity which in turn requires high-energy electrons in order that transit time be held to less than half a period.

The obvious photosurface is the $SbCs_3$ photosurface because of its great sensitivity and because it can be laid down on the glass in a semitransparent layer having a very large light-gathering power. For electron focusing purposes the front surface of the tube was ground to the shape of a sphere of radius of curvature 2 in. Contact to the $\frac{1}{2}$-in. diameter cathode was made by means of a layer of platinum laid down on the glass as shown. The electrode following the cathode was a sheet of molybdenum formed to be a sphere concentric with the cathode.

[14] The large degree of polarization in a magnetic field has an interesting probable explanation. Electrons moving along the magnetic field are quickly swept out of the discharge. Electrons moving at right angles to H are confined to tight spirals. The plasma can be expected, then, to be made up predominantly of electrons whose velocity is normal to H. These can be expected to excite radiation polarized normal to H, the σ lines. These are observed to predominate.

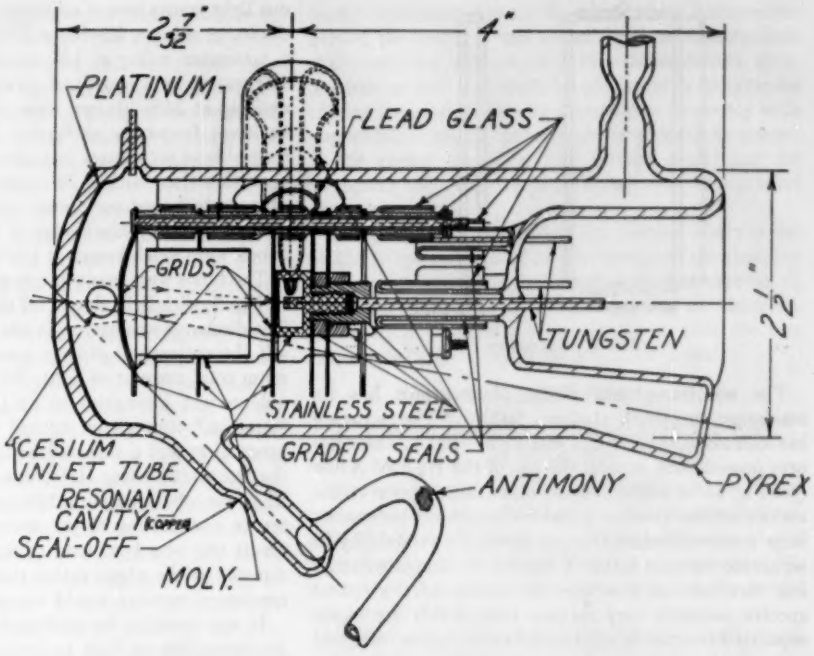


FIG. 4. Photoelectric mixing tube (photomixer).

A grid strung across the ¾-in. diameter hole was woven of 0.0005-in. diameter tungsten wire spaced 0.020 in. between wires. The purpose of the cylindrical electrode is to permit slight adjustments in focal length, and of the flat plate to shield the cavity from electrons which originate elsewhere than at the photosurface and to make possible a focusing prior to examination of the microwave power delivered to the cavity. The entire electrode structure except for the cathode is assembled, as shown in Fig. 4, on three columns of Pb glass beads held together with platinum tie rods, and carried on a single 0.100 in. diameter tungsten lead.

The requirement of high voltage, in order to minimize transit time, and the use of Cs in the photosurface manufacture, almost proved incompatible. After the admission of Cs it becomes impossible to hold high voltage across Pyrex. The Cs appears to form a conducting layer which does not leave the glass at temperatures up to 150°C, the highest temperature to which the tube could be subjected after the photosurface was laid down. However the use of lead glass[15] in strategic locations makes it possible to hold high voltages. Each lead-in had a graded seal and a ½-in. length of lead glass and the support structure was made of lead glass beads held together by a tie rod made of platinum to match the coefficient of expansion of the glass.

Serious technical difficulties associated with the making of seals for bringing energy out of the cavity were eliminated by folding the envelope down into a hole in the side of the cavity, so that a pick-up loop on the end of a coaxial line could be inserted externally. The portion of the glass inside the cavity was a very thin-walled bubble and did not load the cavity measurably. A low-loss glass had been planned for this application but Pyrex was found satisfactory. Here, again, to prevent electrical leakage a lead glass shield had to be used. This method of withdrawing energy from the cavity not only eliminated the problem of difficult vacuum seals but provided the very great advantage of permitting external adjustment of the coupling to the cavity.

To correct for the effects of stray magnetic fields and misalignments, the tube was mounted between two pairs of Helmholz coils so that a variable field in any direction perpendicular to the electron beam could easily be generated.

C. The Light Source

Equation (4) establishes as a figure of merit for a light source photosurface combination $\lambda^2 I/\delta A\Omega$ or $\lambda^2 g\mathcal{J}/\delta$, where g is the photoelectric sensitivity and $\mathcal{J}$ the intensity of the spectral line used. It has been found that the quantity $\mathcal{J}/\delta$ increased with $\mathcal{J}$ and that it is desirable to make the light source as bright as possible, consistent with the condition (1). Although it is not necessarily so, under certain circumstances δ will vary as $1/\lambda$. In that case the figure of merit becomes $\lambda^3 g\mathcal{J}$.

[15] V. K. Zworykin and E. G. Ramberg, *Photoelectricity* (John Wiley and Sons, Inc., New York, 1949), p. 88.

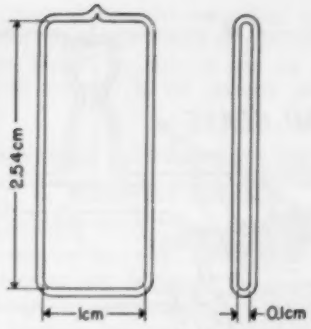


FIG. 5. Light source.

The semitransparent $SbCs_3$ photosurface has its maximum sensitivity at about 5000 A.[16] The factor $\lambda^2 g$ has its maximum at 5300 A and $\lambda^3 g$ at 5400 A. These figures immediately suggest the use of the Hg 5461 A line (see Fig. 5). In addition many other factors point to the use of this line. Mercury is an excellent choice because its large mass minimizes Doppler effect, its availability in separated isotopes makes it possible to eliminate hyperfine structure as a source of broadening, its sparse spectra contains very intense lines which are easily separated from each other and its convenient form and vapor pressure makes its use in discharge tubes extremely convenient. The green line not only occurs close to the optimum wavelength but is very brilliant and there is a great practical convenience in working with the wavelength which has become the standard for retardation plates, nonreflecting coatings, etc. There are two drawbacks to this line. The ease with which the line reverses prevents the use of deep columns at high intensity; and the Zeeman spectra for this line, shown in Fig. 6, are such that only 60 percent of the energy in the σ components is effective in producing beats. It is not inconceivable that a careful examination of the lines λ4046 or λ4358 might make them intrinsically somewhat better than λ5461 but the difficulty of separating the blue and violet lines with filters of sufficient transmissivity rules strongly in favor of the green line. The filter used was made up of 3 mm of Jena glass BG 20 and 2 mm of Jena Glass GG 11 "A" which passed 97 percent of the green line, excluding the effect of surface reflections, while effectively eliminating all of the other mercury lines. To minimize surface reflections the filter components and a polaroid were cemented together with Canada balsam.

In line with recent light source developments,[17–20] our light source was a microwave-excited electrodeless discharge using a Raytheon Microtherm CDM-2 with a maximum rating of 125 watts (but measuring only 100 watts) as a source of power. The advantages of exciting at 2450 Mc/sec were very great as compared to lower frequency excitation. With this supply we were able to get greater intensity and much more stable operation than with a 70-megacycle 2-kilowatt generator we had used previously; and the lifetimes of the tubes, limited by blackening of the walls of the quartz tubes, were much longer at the higher frequencies.[21]

The tubes were made of quartz and shaped as shown in Fig. 5, with light drawn off through the broad faces. The discharge was narrow in one dimension to minimize self-absorption and of large area to produce the maximum total amount of light. With this source we were able to get intensities of 6×10^{-4} watt per cm^2 per steradian[22] through an optical system containing the green filter and a polaroid under conditions for which the line width, using Hg^{202}, was approximately 10^9 cps. This was obtained with 100 percent of power to the source and cooling of the source by directing jets of air at the two 1-in. long edges. The cooling air was directed at the edges, rather than the faces so that the condensed mercury would not obscure the light.

It was possible, by cutting down on the cooling, to get intensities as high as 2.4×10^{-2} watt per cm^2 per steradian through the filter and polaroid, but under circumstances in which the line widths were too great for this experiment.

D. Circuits

The essential circuitry is shown in block form in Fig. 3. The power delivered to the cavity by the electrons is withdrawn by a coupling loop on the end of

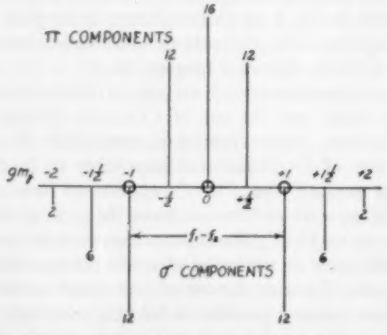


FIG. 6. Zeeman pattern for the 5461 A line of Hg, a $^1S_1 \rightarrow {}^3P_2$ transition. π lines, polarized with the magnetic field, are drawn above and σ lines below. Heights of lines represent relative intensities and the circles indicate the positions in the normal triplet.

[15] V. K. Zworykin and E. G. Ramberg (reference 15), p. 98.
[17] W. F. Meggers and F. O. Westfall, J. Research Natl. Bur. Standards 44, 447 (1950).
[18] Kerr and Des Lattes, Office of Naval Research Technical Report No. 10, Contract Nonr, 248, T. O. 8 (unpublished).
[19] Zelakoff, Wykoff, Aschenhand, and Loomis, J. Opt. Soc. Am. 39, 12 (1949).
[20] E. Jacobsen and G. R. Harrison, J. Opt. Soc. Am. 39, 1054 (1949).

[21] The light source is described much more fully in a separate paper being prepared for another journal.
[22] Intensities quoted are values estimated from the photocurrent from a 929 and the handbook value of photosensitivity at λ5461.

a coaxial line coupling into a waveguide which feeds the power to a magic T. In the T this power is mixed with local oscillator power at a pair of crystals in a technique now common in microwave detection. To avoid the difficult transformation problems encountered in changing a balanced signal from the crystals to a one-sided input for the intermediate frequency (i.f.), the crystals are Microwave Associates IN23CMR, a matched pair of crystals of opposite polarity with respect to the cartridge. This retains the advantage of balanced detection, i.e., the cancellation of local oscillator noise, and simplifies the problem of an appropriate input transformer for the i.f. amplifier. Since the transformer is one of the most important components in achieving a low-noise, wide-band amplifier, this is an important feature of the detector. The i.f. amplifier is a 30-megacycle amplifier with a noise figure of 1.0 db and a band width of 7 Mc/sec with a gain of about 80 db. The i.f. detector output is fed into a 4-cycle wide 45-cps amplifier whose output was rectified by the phase detector which was a mechanical switch made up of four 90° segments, rotating with the half-wave plate which produced the signal modulation, and three brushes at 45° intervals connected as shown in Fig. 4. The phasing was adjusted to produce a maximum output for a signal of the proper phase by rotating polaroid B. The rectified signal was passed through a low-pass RC filter with an RC of 250 seconds, which sets the band pass of the 45-cps detecting circuitry at approximately 1/250 cps. The filter output is fed to a low frequency (0.1-cps) periodic chopper at the input of the balanced dc amplifier which eliminates dc amplifier drift as a source of spurious signal. The effect of the chopper is to discriminate between signals originating prior to the dc amplifier and dc amplifier drift in a manner which is made clear in the photographs of Fig. 7. The positive and negative portions of the chopper cycle are unequal in length to discriminate positive from negative signals. This is particularly important when working near noise level, as in trace 3, Fig. 7, in order to be able to observe, when frequent reversals occur, whether the trace is not more often in one direction than another.

The system has inherent in it a method of checking to see whether an apparent signal originates anywhere else in the circuitry. A rotation of polaroid B by 90° should reverse the sign of a signal which originates in the light source, but leave unaffected a signal originating between the photosurface and the chopper.

Because of klystron drift a device for maintaining the local oscillator frequency at 30 Mc/sec from the cavity frequency is necessary. A 120-cps frequency-modulated signal is reflected from the cavity as shown in Fig. 3. Far from the cavity resonance, the signal displayed on the oscilloscope trace mirrors the i.f. response, as shown by the dotted curve on the oscilloscope. When the local oscillator is properly set with respect to the cavity, the oscilloscope trace shows a dip whose width is characteristic of the cavity Q and whose depth is characteristic of the coupling to the cavity. This simple circuit, then, provides us with a guide to i.f. alignment, a monitor for holding the local oscillator at the correct frequency with respect to the cavity, and a guide for proper adjustment of the cavity coupling. The 120-cps signal and its harmonics are unable to pass through the narrow-band 45-cps amplifier.

The polarization compensation monitor chain is fed by the 45 cps component of the current to the resonant cavity, and is essentially identical to that portion of the main detector chain following the i.f. detector. However, it contains a phase shifter so that the two chains can be set to amplify exactly in phase.

IV. CALCULATIONS

A. Precise Calculations for an Arbitrary Area of Coherence

Because some interesting conclusions can be drawn from an accurate comparison of the observed signal with the theoretical signal strength, and because the detecting system is one which has a wide application, it seems pertinent to present an outline and the results of a precise calculation.

If two spectral lines of equal intensity and equal width, and with Gaussian shapes, are split into Fourier components which are added and squared, representing the action of the photosurface, that part of the current per unit area which arises from differences in frequencies in the two lines is

$$i=[(2\ln 2)/\pi]^{\frac{1}{2}}(i_{d\epsilon}/\sqrt{\delta})\sum_k \exp[-(\ln 2)(\omega_0-\omega_k)^2/\delta^2] \times\cos(\omega_k t+\phi_k)(\Delta\omega)^{\frac{1}{2}}, \quad (6)$$

where ω_0 is the difference in frequency between line centers, δ is the line width, $i_{d\epsilon}$ is the photocurrent per unit area, ϕ_k is a random phase and $\Delta\omega$ is the band width of a single component of the sum. (All frequencies are here expressed in radians/sec.) The current is proportional to cathode area only over a small area $\gamma\lambda^2/\Omega$, referred to as the area of coherence, where Ω is the solid angular spread in the light striking the cathode, λ is the wavelength, and γ is a numerical constant of the order of one.[23] Since the beats from separate areas are randomly phased, the rms beat current from the entire cathode is proportional to the square root of the number of such areas in the cathode surface. This leads to a mean square current, in the frequency range ω to $\omega+\Delta\omega$, including here the effect of the modulation at a frequency ω_M,

$$\langle I^2\rangle=\frac{1}{2}\left(\frac{\ln 2}{2\pi}\right)^{\frac{1}{2}}\frac{\alpha_2^2 I_\epsilon^2\gamma\lambda^2}{\delta\alpha_1 A\Omega} \times\exp\left[-\frac{2(\ln 2)(\omega_0-\omega)^2}{\delta^2}\right](\Delta\omega)\cos\omega_M t, \quad (7)$$

[23] This point is amplified in Sec. IVB.

where I_c is the dc current to the cavity, α_1 is the fraction of the photocathode area from which electrons enter the microwave cavity and α_2 is the fraction of the photoelectrons due to light from the two lines being mixed.

Accompanying this current is the unavoidable shot noise given by

$$\langle I_n^2\rangle = eI_c\Delta\omega/\pi, \tag{8}$$

where e is the electron charge. The currents given by Eqs. (7) and (8) deliver power to the cavity according to

$$P=\frac{\langle I^2\rangle R}{1+4Q^2(\omega-\omega_c)^2/\omega_c^2} \tag{9}$$

where R, Q, and ω_c are the shunt resistance, Q-factor and resonant frequency of the cavity, leading to a power per unit frequency range into the microwave receiver, including receiver noise referred to its input,

$$\frac{dP_1}{d\omega}=\frac{1}{1+4Q^2(\omega-\omega_c)^2/\omega_c^2}+K_F + \frac{K_B\cos\omega_M t}{1+4Q^2(\omega-\omega_c)^2/\omega_c^2}, \tag{10}$$

where

$$K_B=\frac{1}{2}\left(\frac{\pi\ln 2}{2}\right)^{\frac{1}{2}}\frac{\gamma\lambda^2}{e}\frac{I_c\alpha_1^2}{\delta\alpha_1 A\Omega} \tag{11}$$

is the signal-to-shot-noise ratio at the cavity and

$$K_F=FkT(1+\beta)/2eI_cR\beta \tag{12}$$

is the ratio of the contribution of receiver noise to that of shot noise at the central frequency. F is the noise figure of the microwave receiver and β, the coupling coefficient, is the ratio of power withdrawn from the cavity to power dissipated in the cavity. Q and R are both inversely proportional to $(1+\beta)$.

The effect of the superheterodyne detection and i.f. amplification is to lower all frequencies and multiply the power by $G_1(\omega)$, the i.f. power gain, leading to

$$\frac{dP_2}{d\omega}=G_1(\omega)\left\{\frac{1}{1+4Q^2(\omega-\omega_{i.f.})^2/\omega_c^2}+K_F + \frac{K_B\cos\omega_M t}{1+4Q^2(\omega-\omega_{i.f.})^2/\omega_c^2}\right\}. \tag{13}$$

An expression for the output of the detector[24] following the i.f. may be obtained by repeating the process which led to Eq. (6), i.e., representing the power P_2 as the square of the voltage expressed in terms of the Fourier components. If we retain only terms $\omega\approx\omega_M=2\pi\times 45$, we get for the input to the 45-cps amplifier:

$$V=2Y\sum_j\cos(\omega_j t+\phi_j)(\Delta\omega)^{\frac{1}{2}}+K_BX\cos\omega_M t, \tag{14}$$

where

$$X=\int_0^\infty\frac{G_1(\omega)d\omega}{1+4Q^2(\omega-\omega_{i.f.})^2/\omega_c^2}, \tag{15}$$

and

$$Y=\left\{\int_0^\infty G_1^2(\omega)\left[K_F+\frac{1}{1+4Q^2(\omega-\omega_{i.f.})^2/\omega_c^2}\right]^2 d\omega\right\}^{\frac{1}{2}}. \tag{16}$$

The effect of the phase detector plus the low-pass filter is to give, finally, a signal-to-rms-noise ratio[25] at the indicator:

$$S/N=K_BX/2YZ, \tag{17}$$

where

$$Z=\left\{\int_0^\infty[G_2(\omega)/G_2(0)]^2 d\omega\right\}^{\frac{1}{2}}, \tag{18}$$

$G_2(\omega)$ being the voltage gain of the low-pass filter.

In going from Eq. (13) to Eq. (14) we assumed, above, a square-law detector. Our detector, as operated, is more nearly linear than square-law but, according to Selove,[26] this will not alter Eq. (17).

For the low-pass filter shown in Fig. 3,

$$S/N=(RC/2\pi)^{\frac{1}{2}}K_BX/Y, \tag{19}$$

where R is the parallel impedance of the filter resistance and dc amplifier input.

For a flat i.f. response over a band width B,

$$X=B(\tan^{-1}s)/s, \tag{20}$$

where $s=QB/\omega_c$ is the ratio of i.f. to cavity width, and

$$Y=B\left[\frac{\tan^{-1}s}{s}\left(\frac{1}{2}+2K_F\right)+\frac{1}{2}\frac{1}{1+s^2}+K_F^2\right]^{\frac{1}{2}}, \tag{21}$$

so that

$$\frac{S}{N}=\frac{K_B(RCB')^{\frac{1}{2}}(\tan^{-1}s)/s}{\left[\frac{\tan^{-1}s}{s}\left(\frac{1}{2}+2K_F\right)+\frac{1}{2}\frac{1}{1+s^2}+K_F^2\right]^{\frac{1}{2}}}, \tag{22}$$

where $B'=B/2\pi$ is the i.f. band width in cycles per second. For $K_F=0$ and $s\ll 1$, this reduces to $S/N=(RCB')^{\frac{1}{2}}K_B$ in confirmation of Eq. (5).

B. Area of Coherence

There remains to be calculated the constant γ, giving the ratio of the effective area of coherence to λ^2/Ω. It is plain that the area of coherence must be of the order of the size of a diffraction image from the following considerations: Over a diffraction pattern a large phase shift occurs but, since the diffraction pattern is essentially identical for two frequencies very close together, the relative phase shift and therefore the phase of the

[24] In ordinary microwave circuitry this would be the second detector. In this application it is the third, the first being the photosurface.

[25] This result is easily seen to be quite similar to that of R. H. Dicke (reference 12) as expressed in his Eq. (20).

[26] W. Selove, Rev. Sci. Instr. **25**, 120 (1954).

beats remains constant over a diffraction pattern, if the two frequencies originated at a single source point. Of course, the area of a single diffraction pattern contains many overlapping patterns but it is still true that the beats will not be completely random until we get to two image points separated by the size of the diffraction image. The area of the central image is $3.67\lambda^2/\Omega$ so that the effective area of coherence is certainly less than this, the intensity in the secondary diffraction rings being very small.

Gerjuoy[27] has made a calculation of the signal strength based on an integration over the entire cathode using an expression for the field intensity which considers the light as spread uniformly over an area A with a solid angular spread Ω, at each point, and derived an expression for the intensity which an appropriate comparison with the derivation of the preceding section yields a value of $\gamma=1$.

Another, and very interesting, approach to γ may be made through a work of Landé.[28] He derives an expression for the number of degrees of freedom in a beam of light acting for a specified time, which can be converted, for a narrow band of frequencies, into an expression for the number of independent oscillators required to duplicate a light beam of given area and angular spread. If this number of oscillators is equated to the number of areas of coherence, we are again led to $\gamma=1$.

V. EXPERIMENTAL RESULTS

Several recording meter traces are shown in Fig. 7. The signal size is gotten from a comparison of trace 1 with trace 3, corresponding to a rotation of the polaroid B, Fig. 3, through 90°. The signal does not merely change sign in this process, as it should, indicating that some signal originates in the circuitry and this is borne out by trace 2 taken for zero light intensity. The source of the spurious signal apparently lies in a mechanical coupling between the rotating $\lambda/2$ plate and the microwave circuitry. At least it is true that pains taken to isolate these and to make the rotation as quiet as possible reduced this to the low level shown from a very much higher one. The signal obtained in this manner, divided by the rms deviation from the mean was taken for several magnetic field variations and is shown plotted in Fig. 8. The lengths of the lines drawn through each point are the probable error in the results and are less than the rms values of the deviations out of the dc amplifier because the trace was observed for a time long compared to RC for the low-pass filter. Because of darkening of the light source tubes, and other effects, it was not possible to duplicate the conditions for each point and the signal heights have all been corrected to 3.85 μa.

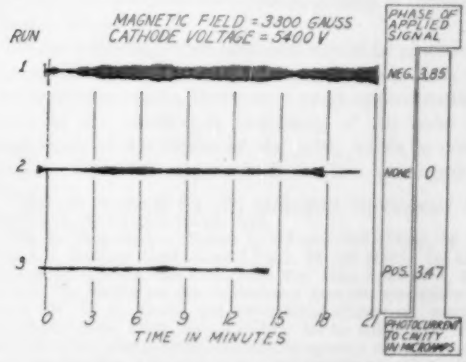


FIG. 7. Photograph of meter traces.

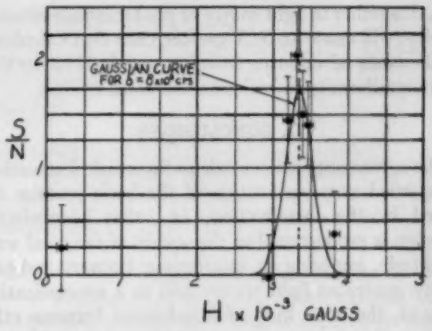


FIG. 8. Final signal-to-noise ratio vs Zeeman field. Cavity current = 3.85 μa. Accelerator voltage = 5400 v.

The relative contribution of circuit noise and shot noise is contained in Eq. (22) and a measurement of the variation in total noise with cavity current yielded, through this equation, a value of $K_F=0.206$ at $I_c=3.85$ μa. [This value in Eq. (12) yields a noise figure for the microwave receiver of approximately 8, which is consistent with that expected from the measured i.f. noise figures and checks the value roughly measured directly.]

For the calculation of K_B by Eq. (11), $\delta/2\pi$ was taken as 8×10^8 cps. This value is obtained from Fig. 7 if it is realized that the width, at half-maximum, of the best spectrum is $\sqrt{2}$ times the spectral line width. The fraction of photocurrent which entered the cavity was $\alpha_1=0.6$, and for the spectral line used (see Fig. 6) $\alpha_2=0.6$. The factor $A\Omega$ was approximately 0.17, leading to $K_B=2.63\times10^{-6}$.

When these values of K_B and K_F together with $RC=240$ sec, $B'=6.8\times10^6$ cps and $s=1$ are used, Eq. (23) leads to

$$(S/N)_{\text{calculated}}=0.83. \tag{23}$$

This is one-half of the observed S/N shown in Fig. 8, close enough to be regarded as good agreement. Furthermore, a spatial variation in emission over the cathode

[27] E. Gerjuoy, University of Pittsburgh (unpublished).
[28] A. Landé, *Handbuch der Physik* (Verlag Julius Springer, Berlin, 1928), Vol. 20, p. 453.

area, due either to light source or photocathode inhomogeneity, will cause an S/N greater than that calculated on the basis of uniform emission and it is likely that the nonuniformity in emission was considerable.

VI. CONCLUSIONS

The agreement between calculations and observations is regarded as a verification of the basic premise discussed in the introduction, i.e., that photoelectric emission is proportional to the square of the total wave amplitude, implying an interference between independently generated light waves, and as a generalization, perhaps, the same kind of interference between other waves such as quantum mechanical wave functions. An alternative to this conclusion would be that there existed some coherence between components of a Zeeman spectrum. However, from a quantum-mechanical point of view two Zeeman lines originate in two distinct sets of atoms with different orientations in the magnetic field and are as independent as if they originated in different atomic species. Furthermore, it seems unlikely that an effect due to a completely unrelated cause would produce so nearly the same size signal as the one we calculated.

A secondary conclusion is that any delay between photon absorption and electron emission must be significantly less than 10^{-10} second, since a decay time equal to the beat period would decrease the signal by a factor of 6.4. This time is so short compared to the lifetimes of allowed transitions in the visible that it becomes significant. For example, the possibility that an electron could be raised to some metastable state from which it might decay to the free electron state is ruled out at least for the antimony-cesium photosurface used in this experiment.

ACKNOWLEDGMENTS

We wish to thank the Office of Naval Research for their financial support and loan of valuable equipment during a portion of this experimentation. We wish to thank Dr. Meggers and Dr. Westfall of the National Bureau of Standards for filling light-source tubes with Hg^{198}, even though in our final runs Hg^{202}, which was kindly loaned by the U. S. Atomic Energy Commission, was used. We are especially grateful to Mr. H. M. Joseph, now with Lear Inc., Research and Development Laboratories, who worked many hours with us, without remuneration, to help achieve a very low noise i.f. amplifier essential to the success of the experiment. Dr. Gordon Locher, of Western Radiation Laboratory, also supplied, without charge, considerable help in the manufacture of the photoelectric mixing tube. We wish to acknowledge also the important role of the extreme glassblowing skill of Mr. Greiner in the photomixer manufacture. We are grateful to Dr. A. J. Allen, Director of the University of Pittsburgh Radiation Laboratory and his staff for their gracious help in the preparation of the manuscript during the period when one of us, A. T. Forrester, was associated with that group.

Above all we wish to acknowledge that the idea was originally conceived by W. E. Parkins and A. T. Forrester and given credence through discussions and calculations in which E. Gerjuoy participated as well. We are also pleased to acknowledge the stimulation of numerous discussions with E. Gerjuoy throughout the progress of this work.

Traveling Density Waves in Positive Columns

SATOSI WATANABE AND NORMAN L. OLESON
United States Naval Postgraduate School, Monterey, California

(Received May 31, 1955)

The continuity equations for positive ions and for electrons coupled to each other by the Coulomb force and the effect of ionizing collisions have a solution representing traveling density waves whose frequencies are widely different from the usual plasma oscillations.

THE aim of this paper is to show that there can exist, in the positive column, traveling waves of ion density and electron density whose physical nature is different from the so-called plasma or plasma-ion oscillations.[1] It is premature to identify these waves with the moving striations, but a further elaboration of the theory accompanied by an adequate consideration of modes of excitation of the waves may perhaps provide a theoretical basis for explaining moving striations. The present paper differs from the previous papers by various authors which tried to explain positive striations with the help of diffusion equations, for they considered only the electron density or assumed equality of the electron and ion densities.[2]

The starting equations are diffusion equations for positive and negative ions. Since the observed wavelengths of striations are larger than the mean free paths and since the observed frequencies are smaller than the reciprocal of the time necessary to establish local equilibrium, the use of the diffusion equations is justifiable. These two diffusion equations are coupled to each other through two physical phenomena. First, the electrostatic field generated by positive space charges reacts on negative charges, and vice versa. Second, both positive and negative ions are created by collisions of electrons with neutral molecules, the density of which is much larger than the ion and electron densities. The coupling between the two ion densities under an external electrostatic field gives rise to the possibility of traveling density waves.

In the following, more emphasis should be placed on the possibility of the existence of such waves than on the numerical results. However, a rough approximation gives as the wavelength something of the order of magnitude of the radius of the tube, which is comparable with the observed data. The propagation velocity of waves in the first approximation is of the order of magnitude of the positive ion mobility times the impressed electric field. This is apparently smaller than the observed velocity of moving striations.[3] However, a qualitative discussion at the end of the paper will show that a further refinement of the theory may probably bring the propagation velocity to the neighborhood of the sound velocity of the neutral gas in the tube.

The basic equations are:

$$\begin{aligned}(\partial n^-/\partial t)+\operatorname{div}(n^-\mathbf{v}^-)-zn^-=0,\\ (\partial n^+/\partial t)+\operatorname{div}(n^+\mathbf{v}^+)-zn^-=0,\end{aligned} \tag{1}$$

with

$$\begin{aligned}\mathbf{v}^-=-K^-\mathbf{E}-(D^-/n^-)\operatorname{grad}n^-,\\ \mathbf{v}^+=+K^+\mathbf{E}-(D^+/n^+)\operatorname{grad}n^+,\end{aligned} \tag{2}$$

where the symbols have the following meanings:

n^+, n^-: positive and negative ion densities, respectively,
$\mathbf{v}^+$, $\mathbf{v}^-$: positive and negative drift velocities,
K^+, K^-: positive and negative mobilities,
D^+, D^-: positive and negative diffusion coefficients,
z: number of ion pairs created per unit time per one electron.

The assumption that z is constant may be justified if the temperature variations can be considered very small. The electric field $\mathbf{E}$ consists of two parts: $\mathbf{E}_0$ and $\mathbf{E}'$, where $\mathbf{E}_0$ is the impressed electric field and $\mathbf{E}'$ the electric field generated by space charges.

$$\begin{aligned}\mathbf{E}=\mathbf{E}_0+\mathbf{E}',\\ \operatorname{div}\mathbf{E}=\operatorname{div}\mathbf{E}'=4\pi e(n^+-n^-),\end{aligned} \tag{3}$$

where e is the electronic charge. We consider only singly charged ions and assume that the tube has a circular cross section and ignore the effect of the both ends of the positive column.

We propose now to solve these equations by assuming that the actual solution is a superposition of a steady,

[1] This can be seen in Eq. (19) which gives the frequency depending on the external electric field.

[2] M. T. Druyvesteyn, Physica **1**, 273 and 1003 (1934); M. F. Shirokov, Doklady Akad. Nauk S.S.S.R. **89**, 837 (1953). See also Armstrong, Emeleus, and Neill, Proc. Roy. Irish Acad. **54-A**, 291 (1951). Our thanks are due to Professor Emeleus who informed us of Dr. V. D. Farris' yet unpublished theory on moving striations. [Proc. Phys. Soc. (London) (to be published).] His work, however, also assumes approximate equality of positive and negative densities. Yoshimoto *et al.*, treat the same problem without this assumption, but their results are essentially different from ours since they do not take creation of ions into consideration. Our results, as can be seen from our Eq. (26), depends crucially on the rate of production of ions. See Yoshimoto, Sato, and Nakao, Repts. Phys. Dept., Okayama Univ. **10**, 15 (1954).

[3] In a tube approximately one centimeter in diameter and containing argon at 12 mm Hg pressure, we have observed positive striations in the glow discharge moving with velocities of 25 meters per second or higher. Using the data of J. A. Hornbeck, Phys. Rev. **84**, 615 (1951), on drift velocities of A^+ in argon, one finds that an average field intensity of 2600 volts per meter would be required to account for the observed velocity. The velocity of sound in argon under the above conditions is 320 meters per second.

uniform solution $n_0^{\pm}$ plus a small nonsteady, nonuniform disturbance $\nu^{\pm}$. The assumption that $\nu^{\pm}$ is small compared with $n_0^{\pm}$ may not be realistic if the theory is to be applied to the actual striations. This assumption thus is a mathematical expediency to linearize the equations and to help discover certain aspects of the complicated physical situation.

$$\begin{aligned} n^+ &= n_0^+ + \nu^+, \quad |\nu^+| \ll n_0^+, \\ n^- &= n_0^- + \nu^-, \quad |\nu^-| \ll n_0^-. \end{aligned} \tag{4}$$

We shall write $\mathcal{E}$ for the electric field generated by the steady, uniform charge densities $n_0^{\pm}$. The word "uniform" here means that the quantities do not depend on the x-coordinate taken along the axis of the tube.

$$\begin{aligned} \partial n_0^+/\partial t = \partial n_0^-/\partial t = \partial \mathcal{E}/\partial t = 0, \\ \partial n_0^+/\partial x = \partial n_0^-/\partial x = \partial \mathcal{E}/\partial x = 0, \\ \mathcal{E}_x = 0. \end{aligned} \tag{5}$$

These n_0^+, n_0^- and $\mathcal{E} = \mathbf{E}'$ must satisfy Eqs. (1), (2) and (3).

The well-known ambipolar diffusion solution[4] may be considered as an approximate solution satisfying these conditions, i.e., Eqs. (1), (2), (3), and (5). This ambipolar diffusion solution is not self-consistent in the sense that the assumption $n_0^- = n_0^+$ leads to $\mathbf{E}' = \mathcal{E} = 0$ according to Eq. (3), while the other assumption which is also used that $\mathbf{v}_r^+ = \mathbf{v}_r^-$ (r meaning radial component) leads, on account of Eq. (2), to the electric field given by

$$\mathcal{E}_r = \frac{D^+ - D^-}{K^+ + K^-} \cdot \frac{1}{n_0} \cdot \frac{dn_0}{dr}, \tag{6}$$

where r is the radial distance from the tube axis.

However, we can expect that a rigorous solution of Eqs. (1), (2), (3), and (5) under adequate boundary conditions is not very different from the ambipolar diffusion solution, in particular in the region not too close to the wall of the tube. We assume specifically that the rigorous solution has, in common with this approximation, the following features. First, in the region which is not very far from the tube axis the densities n_0^+ and n_0^- are almost constant and approximately equal to each other. Second, the rate z of production of ion pairs is roughly given by

$$z = (2.405/R)^2 D_a, \tag{7}$$

where the ambipolar diffusion coefficient D_a is defined by

$$D_a = (D^+K^- + D^-K^+)/(K^+ + K^-) \doteq K^+ \kappa T_e/e, \tag{8}$$

where κ is the Boltzmann constant and T_e is the electron temperature.[4] R in Eq. (7) is the radius of the tube.

Now we substitute Eq. (4) in Eqs. (1) and (2), and simplify the equations with the help of the fact that $n_0^{\pm}$ is the solution of Eqs. (1), (2), and (3), satisfying Eq. (5). The second order term in the disturbance is to be neglected. For instance, in the term $\operatorname{div}(\nu^{\pm}\mathbf{v}^{\pm})$ of Eq. (1), there will appear a term like

$$\operatorname{div}(\nu^{\pm} K^{\pm} \mathbf{e}), \tag{9}$$

where $\mathbf{e}$ is the electric field generated by the disturbance $\nu^{\pm}$, i.e.,

$$\operatorname{div}\mathbf{e} = 4\pi e(\nu^+ - \nu^-). \tag{10}$$

The term (9) is neglected. Then Eq. (1) becomes a set of linear equations for ν^+ and ν^-, which are

$$\begin{aligned} (\partial\nu^-/\partial t) + \operatorname{div}[-K^-(n_0^-\mathbf{e} + \nu^-\mathbf{E}_0 + \nu^-\mathcal{E}) \\ - D^- \operatorname{grad}\nu^-] - z\nu^- = 0. \\ (\partial\nu^+/\partial t) + \operatorname{div}[+K^+(n_0^+\mathbf{e} + \nu^+\mathbf{E}_0 + \nu^+\mathcal{E}) \\ - D^+ \operatorname{grad}\nu^+] - z\nu^- = 0. \end{aligned} \tag{11}$$

where $\mathcal{E}$ and $\mathbf{e}$ are the fields generated respectively by $n_0^{\pm}$ and $\nu^{\pm}$.

As we are interested in the region not too close to the wall, we assume

$$\begin{aligned} \partial\nu^-/\partial r = \partial\nu^+/\partial r = 0, \\ \mathbf{e}_r = 0. \end{aligned} \tag{12}$$

Remembering that $\mathcal{E}_x = 0$, Eq. (5), and that $n_0^{\pm}$ is almost constant in the central region, we can simplify Eq. (11) to the form

$$\begin{aligned} (\partial\nu^-/\partial t) - K^-E_0(\partial\nu^-/\partial x) - D^-(\partial^2\nu^-/\partial x^2) \\ + \gamma_1\nu^- - \gamma_2\nu^+ = 0, \\ (\partial\nu^+/\partial t) + K^+E_0(\partial\nu^+/\partial x) - D^+(\partial^2\nu^+/\partial x^2) \\ + \gamma_3\nu^+ - \gamma_4\nu^- = 0, \end{aligned} \tag{13}$$

with

$$\begin{aligned} \gamma_1 &= 4\pi e K^-(2n_0^- - n_0^+) - z, \\ \gamma_2 &= 4\pi e K^- n_0^-, \\ \gamma_3 &= 4\pi e K^+(2n_0^+ - n_0^-), \\ \gamma_4 &= 4\pi e K^+ n_0^+ + z, \end{aligned} \tag{14}$$

where n_0^+ and n_0^- are approximately equal to each other.

$$n_0^+ = n_0^- (\equiv n_0). \tag{15}$$

The coefficients γ_2 and γ_4 represent the coupling between the two densities.

We now proceed to show that Eq. (13) has a solution which represents traveling waves of ν^+ and ν^- both having the same space-time dependence. Let us insert in Eq. (13) the expression:

$$\begin{aligned} \nu^- &= a^- e^{ikx - i\omega t}, \\ \nu^+ &= a^+ e^{ikx - i\omega t}, \end{aligned} \tag{16}$$

where a^- and a^+ are complex amplitudes. We then obtain

$$\begin{aligned} (-i\omega - iK^-E_0k + D^-k^2 + \gamma_1)a^- - \gamma_2 a^+ = 0. \\ -\gamma_4 a^- + (-i\omega + iK^+E_0k + D^+k^2 + \gamma_3)a^+ = 0. \end{aligned} \tag{17}$$

[4] See for instance J. D. Cobine, *Gaseous Conductors* (McGraw-Hill Book Company, Inc., New York, 1941), p. 236 ff.

Compatibility of these two homogeneous linear equations leads to two real equations:

$$(D^-k^2+\gamma_1)(D^+k^2+\gamma_2)-\gamma_2\gamma_4 +k^2E_0^2(K^++K^-)^2(D^+k^2+\gamma_2)(D^-k^2+\gamma_1)/[(D^+k^2+\gamma_2)+(D^-k^2+\gamma_1)]^2=0, \quad (18)$$

and

$$\omega=kE_0[(K^+D^--K^-D^+)k^2-(K^-\gamma_2-K^+\gamma_1)]/[(D^++D^-)k^2+(\gamma_1+\gamma_2)]. \quad (19)$$

The fact that these Eqs. (18) and (19) have real solutions for k and ω shows the possibility of traveling waves. We can use Eq. (18) to determine k, and then use Eq. (19) to determine the propagation velocity

$$V=\omega/k. \quad (20)$$

In order to obtain an approximate solution of Eq. (18), we introduce two variables

$$\xi=D^-k^2+\gamma_1, \quad \eta=D^+k^2+\gamma_2, \quad (21)$$

and observe the intersection of two straight lines:

$$\eta=\gamma_2+(D^+/D^-)(\xi-\gamma_0), \qquad \eta=\gamma_2-\{(1+g)^2/[g(1+g)^2+gh/\gamma_2]\}(\xi-g\gamma_2), \quad (22)$$

with

$$g=\gamma_2\gamma_4/\gamma_2^2, \quad h=(K^++K^-)^2E_0^2/D^+. \quad (23)$$

The first equation in Eq. (22) is a direct consequence of Eq. (21). The second equation in Eq. (22) is the tangent at $\eta=\gamma_2$ to the curve representing Eq. (18) in the $\xi-\eta$ plane. The substitution of this tangent for the curve is justified by the fact that $D^-\gg D^+$, which means that the straight line: $\eta=\gamma_2+(D^+/D^-)(\xi-\gamma_1)$ cuts the said curve at a point very close to $\xi=\gamma_2\gamma_4/\gamma_2$, $\eta=\gamma_2$. This approximation yields

$$k^2=(\gamma_2 g-\gamma_1)/\{D^-+D^+[g+gh/\gamma_2(1+g)^2]\}. \quad (24)$$

Making use of Eqs. (14) and (15) and noticing that

$$z\ll 4\pi eK^+n, \quad (25)$$

we obtain from Eq. (24):

$$k^2=(K^++K^-)z/(K^+D^-+K^-D^++K^+K^-E_0^2/4\pi ne), \quad (26)$$

or by the help of Eqs. (7) and (8),

$$\lambda=2\pi/k=(2\pi R/2.405)[1+\tfrac{1}{2}a^2/(eE_0/\kappa T_e)^2], \quad (27)$$

where a is the Debye-Hückel radius of the electron:

$$a=(\kappa T_e/4\pi n_0 e^2)^{\frac{1}{2}}.$$

To determine the sign of the propagation velocity V, Eq. (20), let us take the case $E_0>0$. Then, Eq. (19) shows that V is positive (i.e., traveling from anode to cathode) for sufficiently large values of k and negative (i.e., traveling from cathode to anode) for sufficiently small values of k. When k is very large, we obtain from Eq. (19), with the help of the usually satisfied conditions: $(K^+/D^+)\gg(K^-/D^-)$ and $D^-\gg D^+$,

$$V_{k\to\infty}=K^+E_0. \quad (28)$$

V vanishes, when

$$k^2=z/D^-\ll 1\ \text{cm}^{-2}. \quad (29)$$

The amplitude ratio and phase relation between the ion density wave a^+ and the electron density wave a^- can be easily calculated from Eq. (17). Putting

$$a^+/a^-=Ae^{-i\phi}, \quad (A,\phi \text{ real}), \quad (30)$$

we obtain

$$\tan\phi=(K^++K^-)kE_0/[(D^++D^-)k^2+(\gamma_1+\gamma_2)], \quad (31)$$

and

$$A=(1/\gamma_2)(D^-k^2+\gamma_1)(1+\tan^2\phi)^{\frac{1}{2}}. \quad (32)$$

The order of magnitude of the factor $(D^-k^2+\gamma_1)/\gamma_2$ in Eq. (32) is unity, meaning that the positive amplitude and the negative amplitude are of the same order of magnitude unless $\tan\phi$ becomes very large compared with unity.

If we assume $A>0$, $k>0$, $E_0>0$, then we see from Eq. (17) that

$$0<\phi<\pi/2. \quad (33)$$

This means that for a wave propagating from anode to cathode, the phase of the positive density wave is leading that of the negative density wave by an angle between 0 and $\pi/2$. For very large values of k and for very small values of k, this phase difference vanishes, as can be seen from Eq. (31). Between these two limits $\tan\phi$ takes one maximum which is given by

$$(\tan\phi)_{\max}=(K^++K^-)^2E_0/[4(D^++D^-)(\gamma_1+\gamma_2)]. \quad (34)$$

Under the usual discharge conditions, the order of magnitude of Eq. (34) cannot be said to be either extremely large or extremely small.

It should be noted that the actual drift current of ions (as well as of electrons) is composed of three parts: (1) the part due to the external field $n_0^+K^+\mathbf{E}_0$ in the axial direction, (2) the part constituting the ambipolar diffusion in the radial direction, $n_0^+K^+\mathcal{E}-D^+\operatorname{grad}n_0^+$, and (3) the part due to the disturbance. Except for the negligible term given in formula (9), this last part is given by the expression in the brackets following div in Eq. (11). As the radial field $\mathcal{E}$ is negligible in the central region, this part of drift current is in the axial direction, i.e., in the direction of propagation of the wave represented by Eq. (16). Furthermore, the time average of this drift current vanishes since each term in this quantity is oscillating in time. In other words, the traveling wave in question is accompanied by a longitudinal drift current whose time average is zero. In this respect, this wave is a kind of "sound wave," only its mechanism is electrical instead of being dynamical.

The ions, as far as their dynamic properties are concerned, are almost identical with the neutral molecules. Therefore, once the ion density wave described in the

foregoing occurs, a genuine sound wave of neutral molecules will be excited on account of the collisions between ions and molecules. If such a genuine sound wave is excited, the ions will "ride on" this wave, i.e., they will participate on this wave.

The drift velocity determined in Eq. (2) is essentially the terminal velocity in the presence of friction between the neutral molecules and ions due to the relative drift velocity. The calculation given in this paper is based on the assumption that the drift velocity of the neutral molecules is zero. If the neutral molecules perform sound motion, then the friction will be proportional to the difference between the ion drift velocity and the molecule drift velocity accompanying the sound. When these two drift velocities coincide, the friction will disappear and a "resonance" will set in. This situation, which is outside the scope of Eqs. (1) and (2), may explain the fact that the actual propagation velocity of positive striations is found to lie between the value as given by Eq. (28) and the velocity of sound in the gas. We are now undertaking an investigation in this direction.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Helium Film Transfer Rate from 0.14 to 2.19°K*

R. K. WARING, JR.†
Sloane Physics Laboratory, Yale University, New Haven, Connecticut
(Received May 27, 1955)

The transfer rate of the mobile helium film, over Pyrex glass, was measured as a function of temperature in the range from 0.14°K to the lambda point. The bulk of the observations was concerned with film transport due to a thermal gradient but a measurement, using the same apparatus, was made at 1.1°K under a gravitational potential as a comparison. At this temperature it was found that the transport rates, in the two cases, differ by a factor of nearly two, and below 1°K the rate was found to increase with decreasing temperature by about 10%.

I. INTRODUCTION

AMBLER and Kürti[1] have recently measured the rate of transfer by the helium film, under a gravitational potential difference, at temperatures below 1°K. They observed a 30% increase in the rate as the temperature was reduced from 0.85°K to 0.15°K. The only other data, that of Lesinsky and Boorse,[2] whose lowest temperature point was 0.75°K, lend some support to this result. Since all but about one percent of the liquid is already superfluid at 1°K, these observations seem to negate the hypothesis that the temperature dependence of the transfer rate is given by

$$R=(\rho_s/\rho)v_cd,$$

where the product v_cd is a constant approximately equal to $\hbar/2m$. Here R is the flow rate in $cm^3\ cm^{-1}\ sec^{-1}$, v_c is the critical velocity of the superfluid, d is the film thickness, and ρ_s/ρ is the fraction of the liquid which is superfluid.

Liquid helium may also be transferred, by the mobile film, from a position at low temperature to a position at a higher temperature providing that no portion of the surface over which the helium must flow is above the lambda point. Thus thermal film transfer rates may be measured by allowing the film to flow into a region where it evaporates, the rate of vapor efflux then being determined. In view of the results of Ambler and Kürti it appeared desirable to investigate the thermal transfer rates at temperatures below 1°K, and this method offered a convenient means for observing these as well as rates above 1°K.

II. THE APPARATUS

The portion of the apparatus contained in the helium cryostat is schematically represented in Fig. 1. The bulk liquid helium II, which supported the film flow, was contained in a small Pyrex reservoir (*a*), the flow being constricted by a 211-micron i.d. Pyrex capillary (*b*). Except for a baking process to be described, this capillary was untreated. The surface of the liquid in reservoir (*a*) was always farther from the lower end of the Pyrex capillary than the surface tension rise appropriate to a 211 micron tube (~5 mm). This prevented the Pyrex capillary from filling with bulk liquid via the film.[3]

Above this constriction was a glass to Kovar seal (*c*) leading to a 0.014-inch i.d. stainless steel tube (*d*) with a 0.003-inch wall. A tube this fine was used in order to inhibit recondensation, into the reservoir, of helium evaporated from the film. It was necessary, since excessive recondensation would have caused the salt to warm up too rapidly after a demagnetization,

* Assisted by the Office of Ordnance Research, U. S. Army.
† This work formed part of a Ph.D. dissertation.
[1] E. Ambler and N. Kürti, Phil. Mag. **43**, 260 (1952).
[2] L. Lesinsky and H. A. Boorse, Phys. Rev. **87**, 1135 (1952).
[3] Dyba, Lane, and Blakewood, Phys. Rev. **95**, 1365 (1954).

and in addition would have introduced an error into the transfer rate measurements.

Upon leaving the stainless steel capillary the film entered the valve (*f*), which was adapted from a commercial needle valve. This was included, because at temperatures above 1°K the vapor pressure of helium in the reservoir becomes high enough so that large amounts of vapor (apart from that produced by the film) could enter the measuring system and thus create an error. When the valve was "closed" it acted as a superleak preventing the escape of any significant amount of vapor from reservoir (*a*) but offering free passage to the mobile film. A nonopening superleak could not have been used, since it was necessary to pump out the reservoir and constriction thoroughly before cooling them to liquid helium temperatures. Had this not been done, air present would have solidified onto the surface of the constriction giving rise to anomalous results.[4] Further, the valve made possible certain precautions and tests which will be described hereafter. It was actuated from outside the cryostat by means of a thin-walled stainless steel tube (*g*) which was soldered at the top to a brass rod, which in turn passed through an O-ring seal (*h*) in the Dewar cap. The copper tube (*i*) served to bypass the threaded portion of the valve, since this would have acted as a severe constriction in the pumping line.

The film arising in the reservoir was rendered mobile by the thermal gradients in the vicinity of the outer bath level. When the reservoir was below 1°K, there were, of course, other gradients. It passed up to a point at which the temperature was 2.19°K, and here any liquid which had not already evaporated at lower temperatures became vapor. It then passed out of the cryostat through the stainless steel tube (*j*) and into the pumping system. This consisted of a 50-liter/second oil diffusion pump backed by an adequate forepump.

A method which has been used to measure the efflux rate of helium evaporating from the film[5] is to allow it to flow into a previously evacuated bulb and, simultaneously, to measure the time rate of pressure rise in the bulb. For the present experiment, however, instantaneous measurement of the rate was desirable, since the temperature of the helium changed with time in the demagnetization range. The method chosen for this investigation consisted of measuring the fore-pressure in the pumping system as a function of the rate at which helium gas entered the system through a calibrated leak. This leak consisted of an ungreased stopcock with the plug waxed in at top and bottom to prevent the entrance of air. The leak rate through it was varied by changing the pressure on one side of it, the other side leading into the high-vacuum side of the diffusion pump. In order to calibrate the device, a large, accurately known volume was attached to the

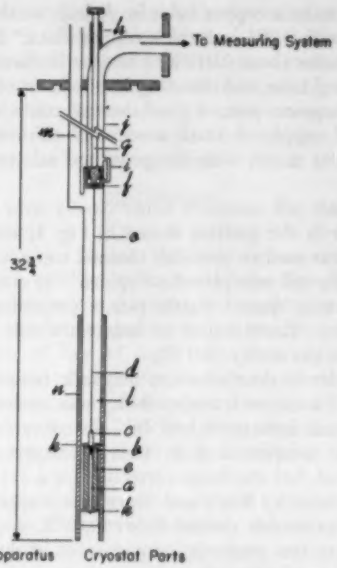


FIG. 1. Schematic diagram of the apparatus contained in the cryostat.

vacuum system; and, with the fore pump closed off and the diffusion pump not operating, the rate of pressure rise in the vacuum system was measured as a function of the pressure of helium on the other side of the leak. A simple calculation converts this to the units in which film transfer rates are expressed. The volume of the part of the vacuum system involved in this measurement was calculated from its dimensions. Since these were small compared with those of the attached container, any errors in the computation resulted in a negligible error in the total volume. The pressure on the leak was measured with a mercury manometer, and the position of the plug in the stopcock was adjusted so that the range of leak rates desired corresponded to a pressure range of about one atmosphere. The pressure in the known volume was measured with a McLeod gauge and did not rise above 100 microns.

In order to calibrate the pumping system, the large known volume was removed, and helium was admitted into the pumping system at known rates, via the leak, with both the diffusion pump and the fore pump operating. The fore pressure was then measured as a function of the known leak rate, and the fore pressures observed during the film flow measurements were then compared with this calibration. A valve between the cryostat and the diffusion pump permitted a pumping system calibration to be made at any time during a run.

Manganous ammonium sulfate was chosen as the paramagnetic salt for this work, since it is known to have a very high specific heat in the neighborhood of 0.15°K. To make the salt pill (*e*) powdered salt was

[4] R. Bowers and K. Mendelssohn, Proc. Phys. Soc. (London) **A63**, 1318 (1950).
[5] H. A. Fairbank and C. T. Lane, Phys. Rev. **76**, 1209 (1949).

pressed onto a copper tube in $\frac{1}{32}$-inch sections interspaced with 0.005-inch-thick copper disks.[6] These had central holes about 0.010 inch smaller in diameter than the central tube, and thus they had to be forced onto it. These measures insured good thermal contact between salt and copper. A small amount of silicone vacuum grease was mixed with the powdered salt to act as a binder.

The salt pill assembly fitted closely over the glass reservoir in the position shown in Fig. 1, and silicone grease was used to establish thermal contact. On each end of the pill was placed a "spider" (*k*) consisting of eight evenly spaced quartz points projecting from a brass ring. These served to isolate the salt from the exchange gas cavity wall (*l*).

In order to determine the magnetic temperature of the salt its magnetic susceptibility was measured using the mutual inductance coil (*n*). No conversion of the magnetic temperatures to Kelvin temperatures was attempted, but the shape correction for a 3:1 ellipsoid, as calculated by Kürti and Simon,[7] was applied. Thus the temperatures quoted below 0.95°K actually approximate the magnetic temperatures of a sphere of the salt used. The measuring coil had two bucking secondaries, and the short ferromagnetic Kovar section (*c*) of the capillary line (*d*) was placed accurately between them. In this position it proved to have a negligible effect upon the mutual inductance. The exchange gas pressure was controlled with a standard vacuum system employing a five liter per second oil diffusion pump. Warm-up times were on the order of two hours. The apparatus shown in Fig. 1 was contained in a liquid helium bath whose temperature was controlled and measured in the usual manner.

RESULTS

The transfer rates observed are plotted against temperature in Fig. 2. The scatter may be partly a result of day-to-day variations in the pumping speed of the measuring system. Two features of the curve appear of particular interest. First, it displays partial agreement with the results of Ambler and Kürti. The increase observed below 1°K is, however, in the neighborhood of 10% rather than 30%, and it saturates at about 0.5°K. It should be emphasized that these measurements were made under entirely different experimental conditions than were those of the above mentioned authors. In the present work the film was rendered mobile by thermal means, and radiation was completely excluded.

The second noteworthy fact is that the rates above 1°K prove to be considerably higher than accepted gravitational rates in the same region. This is in disagreement with the often repeated statement in the literature that critical thermal flow rates are identical

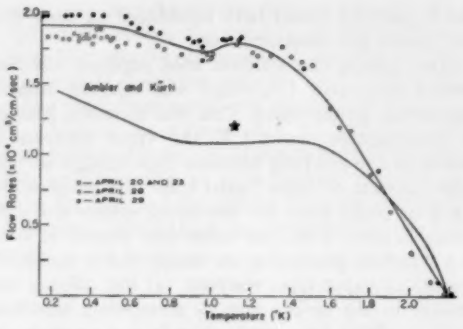


FIG. 2. Dependence of critical film transfer rate upon temperature. The upper curve presents the results of the present work. The star signifies the gravitational flow rate obtained with the same equipment. The lower curve shows the smoothed values obtained by Ambler and Kürti (reference 1).

with the gravitational ones. In the next section we shall present certain arguments to support the validity of the present measurements.

SUPPORTING ARGUMENTS

Extensive precautions were taken in order to avoid contaminating the surface of the constricting Pyrex capillary. The entire vacuum apparatus, including the system for condensing helium into the reservoir, was tested for leaks using a helium mass spectrometer leak detector. Before cooling down the system was pumped out for about 48 hours. A cold trap was used at all times to prevent migration of pump oil. The pressure reached 5×10^{-6} mm as measured with an ion gauge between the cold trap and the diffusion pump.

After cooling down to about 1°K, helium was slowly injected into the system through a charcoal trap held at liquid air temperature. During the process the valve (*f*) was closed in order to prevent the passage into the reservoir of any impurities remaining in the helium. To preclude the possibility of low vapor pressure contaminants which might have been condensed onto the constricting surface during soldering operations on the vacuum system the reservoir and capillary were baked at 300°C for two hours while the system was evacuated through a liquid air trap. That temperature was regarded as high enough, since the only joints made after the glass was attached were soft soldered. Actually, this operation was performed after the runs of April 20–23 (see Fig. 2) and no significant change resulted.

As a final check the reservoir was broken off above the 211-micron constriction, and a gravitational film flow measurement was made with it at 1.1°K. No exceptional precautions were taken to avoid air contamination, but the result (indicated by the star on Fig. 2) falls within the range of the latest reported rates for gravitational flow over Pyrex.[8]

[6] J. C. King and H. A. Fairbank, Phys. Rev. **93**, 21 (1954).
[7] N. Kürti and F. Simon, Phil. Mag. **26**, 849 (1938).
[8] B. Smith and H. A. Boorse, Phys. Rev. **98**, 328 (1955).

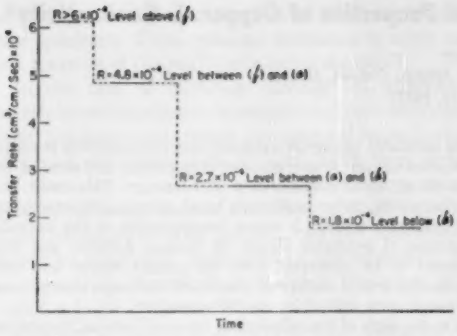

R>6×10⁻⁴ Level above (f)
R=4.8×10⁻⁴ Level between (f) and (a)
R=2.7×10⁻⁴ Level between (a) and (d)
R=1.8×10⁻⁴ Level below (d)
Transfer Rate (cm³/cm²/Sec)×10⁴
Time

FIG. 3. Flow rates for liquid levels above the Pyrex capillaryl At the point (0) there was a constriction in the stainless stee. tube.

In order to insure that the flow was indeed being constricted at the 211-micron Pyrex capillary, enough helium was admitted to bring the liquid level up into the valve. The transfer rate was then measured as a function of time (Fig. 3). Each step corresponds to liquid levels between known constrictions, and the length of each step corresponds to the mass of liquid removed. The latter fact confirms the validity of the method for measuring the rate of vapor efflux, since known amounts of liquid flowed out in times corresponding to the transfer rates as measured. To support the observations below 1°K it was also necessary to demonstrate that the flow was not then being constricted at a point whose temperature was controlled by the bath. This was done by raising the bath from 0.95 to 1.7°K while the reservoir was well below 1°K. No significant change in the transfer rate was observed.

A calculation based upon the extrapolated thermal conductivity of glass,[9] upon the specific heat of liquid helium[10] below 1°K, and upon the quality of thermal contact between salt and copper[11] indicates that thermal equilibrium between the salt and the liquid in reservoir (*a*) should have been reached within a few minutes after demagnetization. Further, the approximate rate of flow of vapor through the stainless steel tube could be determined, with the reservoir (*a*) in thermal contact with the helium bath, by noting the difference between the observed transfer rates with the valve open and those with it closed. This was on the order of one percent for bath pressure of 80 microns. From this one may conclude that the reverse flow occurring with the salt below 1°K was too small to carry an appreciable amount of heat into the reservoir, and thus disturb the temperature equilibrium there.

DISCUSSION

Experiments on the rate of thermal film flow above 1°K fall into two classes. The class represented in the present work includes those in which the film creeps up to a region where it evaporates, the resulting mass rate of flow of vapor being determined and converted to volume rate of flow of liquid. The second class, the isolated beaker experiments, consists of measurements of the volume rate of flow between baths at different temperatures or with different heat inputs and the results cannot be compared with those being reported here.

In addition to the present experiment three others of the first type have been described,[5,12,13] and they all yielded thermal rates lower than those reported here. However, each appears to be open to one or both of the following criticisms. First, in the cases of references 5 and the first experiment reported in reference 13, one would be inclined to question whether all the helium evaporating from the film actually arrived at a point where it could be observed. Second, in the work reported in reference 13 the constricting periphery was not necessarily isothermal throughout its length, and the film flow rates quoted depend upon the assumption that this will have no effect.

On the basis of the evidence so far accumulated it does not appear impossible that a real difference exists between the film flow rates under, respectively, gravitational and thermal impetus.

ACKNOWLEDGMENTS

I should like to express my gratitude to Professor C. T. Lane who suggested this experiment and under whose guidance it was carried out. I am also indebted to Dr. E. Ambler of the National Bureau of Standards for kindly allowing me to examine his Oxford dissertation.

[9] R. Berman, Advances in Physics **2**, 103 (1953).
[10] Hull, Wilkinson, and Wilks, Proc. Phys. Soc. (London) **A64**, 379 (1951).
[11] E. Mendoza (private communication).
[12] B. V. Rollin and F. Simon, Physica **6**, 219 (1939).
[13] E. Ambler and N. Kürti, Proceedings of the International Conference on Low Temperature Physics, Oxford, 1951 (unpublished), p. 70.

Effect of Point Imperfections on the Electrical Properties of Copper. I. Conductivity*

F. J. BLATT
Physics Department, University of Illinois, Urbana, Illinois
(Received April 18, 1955)

Changes in resistivity due to the addition of substitutional impurities and due to the presence of vacancies or interstitials in copper have been evaluated. It was assumed that the free-electron approximation is valid and that Matthiessen's rule holds. Relaxation of the lattice about the imperfections was neglected. The scattering potentials for the various imperfections were derived from the appropriate Hartree fields, and the partial wave method was employed in the evaluation of the scattering cross sections. In each case, the potentials were adjusted until the phase shifts satisfied the Friedel sum rule. The phase-shift calculations were carried out on the ILLIAC (University of Illinois Graduate School High Speed Electronic Digital Computer).

The calculated resistivities due to small concentrations of gallium, germanium, and arsenic in copper, though consistently too large, are in fair agreement with measurements of Linde. The calculated change in resistivity due to interstitials in copper, 1.4 μohm-cm per atomic percent, is essentially the same as that due to an equal concentration of vacancies. This result is at variance with earlier predictions based on annealing experiments on irradiated copper. A recent interpretation of the annealing spectrum of irradiated copper by Cooper, Koehler, and Marx appears to be consistent with the present results and offers agreement with a variety of theoretical and experimental work, although some difficulties remain unresolved.

On the basis of the calculations reported here and comparison with Linde's results for dilute substitutional alloys of copper, it is concluded that the resistivity increase due to a concentration of one atomic percent of Frenkel defects in copper is probably between 1.5 and 2.0 μohm-cm, with vacancies and interstitials contributing about equally to the scattering of conduction electrons.

I. INTRODUCTION

DURING the past few years interest in the effects of nuclear radiation on solid materials has grown considerably. In particular, radiation damage of metals has been the subject of many experiments and much discussion and speculation.[1,2] These investigations are, in a general sense, directed toward an understanding of the properties of imperfections introduced into the metallic lattice as a result of bombardment by fast, massive particles, and also toward a knowledge of the effects which these imperfections have on the physical properties of the substance under examination.

By far the most numerous imperfections which appear as a result of irradiation are vacancies and interstitials, these being created by collisions of the bombarding particle with atoms of the lattice. If the energy of the incident particle is sufficient, the particle will not only be capable of creating single interstitial-vacancy pairs with each collision, but the primaries so formed may in turn eject secondaries from the normal lattice sites. The problem of the total number of interstitial-vacancy pairs to be expected as a result of bombardment by massive charged and uncharged particles of known energy has been treated by Seitz,[3] and more recently by Harrison and Seitz[4] and by Snyder and Neufeld.[5] These more recent investigations indicate that the original results obtained by Seitz probably underestimated slightly the number of Frenkel defects produced by irradiation.

At sufficiently elevated temperatures the vacancies and interstitials created by radiation become mobile, and annealing of radiation damage takes place. The annealing spectrum is of fundamental interest and has been the object of considerable experimental study. The results of such investigations, yielding values for the mobility energy of the imperfection which is believed to be responsible for the observed annealing peak, are of much concern also in studies of self-diffusion[6] and nuclear resonance.[7]

Vancancies and interstitials as well as other point imperfections, such as substitutional impurities, produce measurable changes of many of the physical properties of the metal in which they are present. For example, conductivity,[8] thermoelectric power,[9] elastic constants,[10,11] and rate of self-diffusion,[12–14] are among the numerous properties which exhibit significant changes for relatively small impurity concentrations.

Because electrical measurements can be performed with high precision and relative ease, by far the most numerous data available pertain to changes of the electrical properties, particularly of the conductivity. Point imperfections in otherwise pure metals cause an increase in the resistivity proportional, at low concentrations, to the concentration of the imperfection. If the resistivity change resulting from the presence of one atomic percent of a given imperfection is known, resistivity measurements lead most directly to a

* Supported by the Office of Naval Research.
[1] G. J. Dienes, Ann. Rev. Nuc. Sci. 2, 187 (1953).
[2] J. S. Koehler and F. Seitz, Z. Physik 138, 238 (1954).
[3] F. Seitz, Discussions Faraday Soc. 5, 271 (1949).
[4] W. A. Harrison and F. Seitz, Phys. Rev. 98, 1530(A) (1955).
[5] W. S. Snyder and J. Neufeld, Phys. Rev. 94, 760 (1954).
[6] H. B. Huntington and F. Seitz, Phys. Rev. 61, 315 (1942).
[7] D. F. Holcomb and R. E. Norberg, Phys. Rev. 93, 919 (1954).
[8] J. O. Linde, Ann. Physik 10, 521 (1931); 14, 353 (1932); 15, 219 (1932).
[9] Borelius, Keesom, Johansson, and Linde, Leiden Comm. 206a, 206b (1930).
[10] W. Köster and W. Rauscher, Z. Metallkunde 39, 111 (1948).
[11] G. J. Dienes, Phys. Rev. 86, 228 (1952).
[12] R. E. Hoffman and D. Turnbull, J. Appl. Phys. 23, 1409 (1952).
[13] D. Lazarus, American Society of Metals Meeting, Chicago, October, 1954 (unpublished).
[14] E. Sonder, Phys. Rev. 98, 245(A) (1955).

knowledge of the concentration of that imperfection in a specimen. Thus, residual resistance is often used as a measure of the purity of a metal sample.[15]

In the case of radiation damage the analysis of resistivity measurements is complicated for two reasons. First, bombardment creates two types of imperfections, vacancies and interstitials. Second—even more serious an obstacle to an interpretation of experimental data—, the magnitude of the effect of these imperfections, particularly of interstitials, on the resistivity is somewhat uncertain. Estimates of the resistivity due to one atomic percent of interstitials in copper range from 0.6 to 10 μohm-cm.[16,17]

In order to arrive at a more reliable value for the resistivity increase due to interstitials in copper, a detailed calculation of the cross section for scattering of conduction electrons by an interstitial copper ion was undertaken. The present work is similar to an investigation of Jongenburger[18] concerning the effect on the resistivity of vacancies in copper.

In Sec. II, the method of calculation is briefly described. Section III is devoted to the application of this method to a calculation of the resistivities of dilute solid solutions of gallium, germanium, and arsenic in copper. The results which are obtained and the measured resistivity changes due to these impurities are compared, and it is found that the calculated resistivity changes are consistently too large although they are of the correct magnitude. The method is next applied, in Sec. IV, to the calculation of the resistivity changes due to interstitials in copper. In Sec. V the results of these calculations are discussed and compared with results of recent annealing experiments on irradiated copper. The experiments are examined in the light of the present and previous theoretical as well as experimental work, and a tentative interpretation of the annealing spectrum is advanced which is consistent with many related results, although certain difficulties still remain unresolved. A conclusion summarizes the significant results and speculations.

II. METHOD OF CALCULATION

A. The Scattering Cross Section

The following assumptions, commonly made in calculations of the resistivity due to point imperfections, are used:

(1) The free-electron approximation for conduction electrons is valid.

(2) The scattering potential associated with the imperfection is spherically symmetric.

(3) Matthiessen's rule is valid.

(4) The effect on the resistivity of relaxation of the lattice about the imperfections is negligible.

[15] G. Chaudron, Nature 174, 923 (1954).
[16] D. L. Dexter, Phys. Rev. 87, 768 (1952).
[17] A. W. Overhauser, Phys. Rev. 94, 1551 (1954).
[18] P. Jongenburger, Phys. Rev. 90, 710 (1953); Appl. Sci. Research B3, 237 (1953).

The first three assumptions need no further discussion; the effect of relaxation of the lattice about imperfections had not been considered in detail until Dexter showed that the last of the above assumptions is justified.[16]

The resistivity increment due to a concentration of one atomic percent of a point imperfection is given by[19]

$$\Delta\rho=\frac{\hbar k}{100ne^2}\int[1-\cos\theta]\sigma(\theta)d\omega, \tag{1}$$

where k is the magnitude of the wave vector of the electrons at the Fermi surface, n is the number of free electrons per atom, θ is the angle through which the electron is scattered by the imperfection, $\sigma(\theta)$ is the differential scattering cross section, and $d\omega$ is the element of solid angle.

The differential scattering cross section is calculated by the partial wave method.[20] The application of this method, in which the scattering cross section is given in terms of the phase shifts δ_l, requires the solution of the radial wave equation

$$\frac{d^2u_l}{dr^2}+\left[k^2-U(r)-\frac{l(l+1)}{r^2}\right]u_l=0, \tag{2}$$

where $U(r)=2mV(r)/\hbar^2$, and $V(r)$ is the potential associated with the scattering center.

Since analytic solutions of Eq. (2) can be obtained only for the very simplest potentials, the differential equation usually must be integrated by numerical methods. Generally, the partial wave method is suitable when the only nonvanishing phase shifts are those for small values of l. For the potentials which appear appropriate to scattering by either substitutional impurities or vacancies and interstitials, $\delta_l\simeq 0$ for $l\geqslant 4$ and for energies corresponding to those of conduction electrons in copper. Although it is, therefore, necessary to perform at most four numerical integrations for each potential, a great deal of time is still required if this is to be done with the aid of a desk calculator only. The task would have been all the more tedious because the criterion which, in part, determines the potential that is to be used, involves the phase shifts themselves [see Eq. (5) below]. Fortunately, Dr. R. Rubenstein made available to the writer a program he had written for the solution of the radial wave equation by the ILLIAC,[21] and the calculation could be performed in very short order. The coding was done most expertly by Mrs. M. C. Huse.

Once the phase shifts are known, the differential

[19] N. F. Mott and H. Jones, *Theory of Metals and Alloys* (Oxford University Press, London, 1936).
[20] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949).
[21] R. A. Rubenstein, Ph.D. thesis, University of Illinois, 1954 (unpublished).

scattering cross section is obtained from the expression

$$\sigma(\theta)=\frac{1}{k^2}\left|\sum_{l=0}^{\infty}(2l+1)\exp(i\delta_l)\sin\delta_l P_l(\cos\theta)\right|^2. \quad (3)$$

The resistivity can be evaluated directly in terms of the phase shifts. When Eq. (3) is substituted into the integrand of Eq. (1) the following results[22]:

$$\Delta\rho=\frac{4\pi\hbar}{100ne^2k}\left[\sum_{l=0}^{\infty}(2l+1)\sin^2\delta_l - 2l\sin\delta_{l-1}\sin\delta_l\cos(\delta_{l-1}-\delta_l)\right]. \quad (4)$$

As a partial check on the self-consistency of the potential, one may employ the Friedel sum rule.[23] Friedel has shown that if N is the number of electrons which must be attracted locally to the imperfection to give effective screening, then

$$N=-\frac{2}{\pi}\sum_{l=0}^{\infty}(2l+1)\delta_l. \quad (5)$$

In the case of a vacancy in a monovalent metal $N=-1$, for an interstitial $N=+1$.

B. Choice of Potential

Changes in resistivity of the ideal metals due to addition of small concentrations of substitutional impurities were measured by Linde,[8] and his results were explained by Mott,[24] who calculated the cross section for scattering of conduction electrons by a substitutional impurity in the free-electron approximation. Mott assumed that the impurity could be represented by the screened Coulomb potential of a point charge Ze, where Z is the valence difference between solvent and solute atom, and he made use of the Born approximation in the evaluation of the scattering cross sections. He found that agreement between calculated and experimental results could be obtained only if the screening radius, which appears as a parameter in the expression for the resistivity increase, was chosen to be about half that predicted by the Fermi-Thomas model. However, an important empirical conclusion of Linde, namely that the change of resistivity is proportional to the square of the valence difference, is a direct consequence of Mott's calculation.

Recently, Friedel has redetermined the scattering cross section from a screened Coulomb potential using a partial wave analysis.[25] He obtained reasonably good agreement with Linde's results by requiring that the phase shifts give the correct Friedel sum when substituted into the right-hand side of Eq. (5), N being set equal to the valence difference Z. He also found that under these conditions the appropriate screening radii did not differ greatly from those obtained by the Fermi-Thomas model.

In his calculation of the resistivity due to vacancies in copper, Jongenburger[18] also made use of the free-electron approximation, but he assumed that the removal from the lattice of an ion with its associated potential is equivalent to adding, at that position in the lattice, a potential of opposite sign. He, therefore, chose for his scattering potential the negative of the Hartree potential of a free copper ion,[26] and this potential he then modified to account for the screening by the conduction electrons. Since the vacancy must be neutral, screening is accomplished by effectively displacing one conduction electron from the vicinity of the vacancy. Jongenburger assumed that this screening can be approximated by creating in the electron gas a spherical hole of unit charge and of radius r_s, where r_s is defined by $4\pi r_s^3/3=\Omega$, Ω being the atomic volume. On calculating the phase shifts he found that they nearly satisfied the Friedel sum; substitution of these phase shifts into Eq. (4) gave for the resistivity due to one atomic percent of vacancies in copper

$$\Delta\rho_v=1.25\ \mu\text{ohm-cm}.$$

The scattering potentials for the calculations which will be reported in the following sections were obtained by taking the difference between the Hartree potentials for the case with and without the presence of the imperfection, and then adjusting the screening charge distribution until the phase shifts satisfied Eq. (5). It was found that if, in Jongenburger's calculation of the resistivity due to vacancies, r_s is used as an adjustable parameter and so chosen that Eq. (5) holds exactly,

$$\Delta\rho_v=1.33\ \mu\text{ohm-cm}.$$

It has been pointed out by Friedel[25] that the calculated resistivities do not depend critically on the choice of potential as long as Eq. (5) is satisfied. Thus, for example, the use of a square barrier of extension equal to the radius of one atomic volume may be used to represent a vacancy. If the height of the barrier is adjusted so that the Friedel sum is -1, the resistivity which is obtained is[27]

$$\Delta\rho_v=1.28\ \mu\text{ohm-cm},$$

in agreement with the values quoted above. For calculations of the thermoelectric power, however, the choice of potential appears to be more critical, and, in the case of substitutional arsenic in copper, the potential derived from the Hartree self-consistent fields by the method which has just been outlined gives better agreement with experiment than an equivalent square potential well.[28]

[22] K. Huang, Proc. Phys. Soc. (London) **60**, 161 (1948).
[23] J. Friedel, Phil. Mag. **43**, 153 (1952).
[24] N. F. Mott, Proc. Cambridge Phil. Soc. **32**, 281 (1936).
[25] J. Friedel, Advances in Physics **3**, 446 (1954).
[26] D. R. Hartree, Proc. Roy. Soc. (London) **A157**, 490 (1936).
[27] F. Abelès, Compt. rend. **237**, 796 (1953).
[28] F. J. Blatt (to be published).

III. SUBSTITUTIONAL IMPURITIES IN COPPER

The Hartree potentials for gallium, germanium, and arsenic for various states of ionization are known,[29] and the resistivity increments due to small concentrations of these impurities in copper have been measured by Linde.[6] These substitutional dilute alloys of copper are, therefore, ideally suited to test the validity of this method of determining the resistivity increases due to point imperfections and of the assumptions which have been made.

Since the valences of these impurities exceed that of copper, the scattering potentials will be attractive, and the screening electrons may go into bound states in such potentials. This is likely to be the case for the two 4*s*-electrons of the impurities considered here, and may also occur for all but one of the remaining valence electrons. Indeed, it was found that the phase shifts added to nearly the correct Friedel sum if a scattering potential was used which is the difference in the Hartree potentials of Cu^+ and, for example, As^+. However, use of the Hartree potentials of As^{++} or As^{+++} screened by one, respectively two, negative charges distributed uniformly in one atomic volume, did not give rise to phase shifts which met with the condition imposed by Eq. (5). This is indicated in Table I which also lists the calculated and measured resistivity increases per atomic percent due to these impurities when in solid solution in copper. The calculated resistivities were obtained by making slight adjustments in the scattering potentials so that the Friedel sums were exactly equal to Z. The required changes were made by extending the range of the scattering potential for the case of Ga^+, and reducing the ranges for the cases of Ge^+ and As^+; these adjustments were accomplished by slightly altering the screening charge distributions, leaving the core potentials unchanged.

From the tabulation of the calculated and experimental resistivities it appears that the calculated resistivities, although of the correct magnitude, are consistently larger than the measured. Figure 1 shows the results of the present calculation and the resistivity

TABLE I. Resistivities due to one atomic percent of gallium, germanium, and arsenic in copper. The state of ionization of the impurity refers to the Hartree potential which was employed in arriving at the scattering potential for use in Eq. (2).

Impurity and state of ionization	$(2/\pi)\Sigma_l(2l+1)\delta_l$	$N=Z$	$\Delta\rho$ (calc)	$\Delta\rho$ (meas)
Ga^{+++}	1.0	2.0		
Ga^+	1.86	2.0	2.7	1.4
Ge^{++}	1.87	3.0		
Ge^+	3.2	3.0	5.0	3.75
As^{+++}	2.0	4.0		
As^{++}	2.75	4.0		
As^+	4.25	4.0	8.2	6.8

[29] A. J. Freeman, Phys. Rev. 91, 1410 (1953).

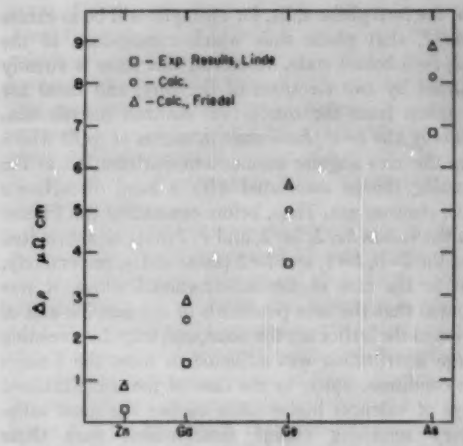


FIG. 1. Resistivity increase due to small concentrations of impurities in copper. Circles show the results of the present calculation; squares show the measured resistivities obtained by Linde; triangles are the calculated values obtained by Friedel.

changes as measured by Linde, as well as those obtained by Friedel using a screened Coulomb potential to represent the impurity. It is seen that the best agreement occurs for large valence differences; this is not particularly surprising in view of the fact that nearly all effects except those attributable to valence difference were neglected. Part of the discrepancy may well be due to neglect of relaxation of the lattice, since, as Dexter has shown,[16] relaxation tends to reduce the resistivity changes. On the whole, agreement between calculated and experimental values is reasonably good. Further justification for the use of scattering potentials derived from the Hartree fields may be the good agreement between calculated and measured thermoelectric power changes due to small concentrations of arsenic in copper.[28] It is also significant that the most satisfactory potentials are those obtained by taking the difference between singly ionized copper and the impurity in the same state of ionization. Apparently, the screening of the impurity is accomplished by electrons whose charge distribution closely resembles that of electrons occupying bound states in the Hartree potential of the free impurity ion.

IV. RESISTIVITY DUE TO INTERSTITIALS IN COPPER

An interstitial copper atom presumably enters the copper lattice in the state Cu^+. Hence, for the scattering of electrons by an interstitial in copper, the scattering potential which is to be used in our approximation is the Hartree potential of Cu^+, suitably screened.

The self-consistent potential of Cu^+ is, of course, sufficiently strong to bind all of the core electrons. In particular, the bound states lying closest to zero energy are the 3*s*, 3*p*, and 3*d* states. In calculating the phase shifts for an electron of energy slightly greater than

zero, the $l=0$ phase shift, for example, will be in excess of $5\pi/2$, that phase shift which corresponds to the $n=3$, $l=0$ bound state. However, this state is already occupied by two electrons of the core, and these are not taken from the conduction electron distribution. It is only the $l=0$ phase shift in excess of $5\pi/2$ which gives the zero angular momentum contribution to the screening charge associated with a local disturbance of the electron gas. Thus, before evaluating the Friedel sum the values $5\pi/2$, $3\pi/2$, and $\pi/2$ must be subtracted from the $l=0$, $l=1$, and $l=2$ phase shifts, respectively.

As in the case of the substitutional alloys, it was assumed that the core potentials of the free ion and of the ion in the lattice are the same, and only the screening charge distribution was adjusted to meet the Friedel sum condition. Since, in the case of the substitutional alloys of valences higher than copper the most satisfactory screening charge distributions were those corresponding to bound states, a screening charge distribution resembling that of a $4s$-electron was used for the interstitial calculation. This is, of course, a somewhat arbitrary choice, but, as anticipated,[25] the results are not very sensitive to these particular details as long as the phase shifts conform to Eq. (5).

Figure 2 shows twice the effective nuclear charges for singly ionized copper,[26] for the assumed screening charge distribution, and for the resulting interstitial potential as functions of radius. For this potential the calculated phase shifts are

$$\delta_0=3\pi+0.258;\quad \delta_1=2\pi+0.516;\quad \delta_2=0.252;\quad \delta_3=0.011.$$

When the phase shifts corresponding to the highest bound states of the core are subtracted from the above values one finds that the Friedel sum is unity. Substitution of the phase shifts into Eq. (4) gives for the resistivity increase due to one atomic percent of interstitials in copper:

$$\Delta\rho_i=1.40\ \mu\text{ohm-cm}.$$

Twice the effective nuclear charges for two different screening distributions are shown in Fig. 3. The solid curve is the same as the corresponding one of Fig. 2; the other two distributions were also adjusted until the phase shifts calculated from the resulting scattering potentials satisfied the Friedel condition. The sharp cut-off distribution (dot-dash curve) leads to the following phase shifts:

$$\delta_0=3\pi+0.32;\ \delta_1=2\pi+0.52;\ \delta_2=0.24;\ \delta_3=0.01;$$

the corresponding resistivity increase per atomic percent is

$$\Delta\rho_i=1.29\ \mu\text{ohm-cm}.$$

The scattering potential obtained by using the long-tailed distribution (dotted curve) gives rise to phase shifts:

$$\delta_0=3\pi+0.22;\ \delta_1=2\pi+0.50;\ \delta_2=0.27;\ \delta_3=0.01.$$

Substitution of these phase shifts into Eq. (4) leads to a resistivity change:

$$\Delta\rho_i=1.41\ \mu\text{ohm-cm per atomic percent}.$$

It is seen that although the assumed screening charge distributions for these three cases differ considerably (the effective nuclear charges are drawn to logarithmic scale in Fig. 3) the change in the calculated resistivity increase is not very significant.

Thus, the calculated increase in resistivity due to the presence of interstitials in copper is about 1.4 μohm-cm per atomic percent, and this is practically the same as the resistivity due to an equal concentration of vacancies.

To the extent to which it is valid to extrapolate the results of the previous section to the effect of Frenkel defects on the resistivity of copper, it is reasonable to presume that the calculated values reported here and also by Jongenburger overestimate the true resistivity increase. A reasonable estimate for the change in resistivity due to vacancy-interstitial pairs in copper is probably 1.5–2.0 μohm-cm per atomic percent, with each type of lattice defect contributing about equally to the total. These conclusions are in good agreement with earlier estimates of Dexter.[16]

V. DISCUSSION OF RESULTS

In recent years, a number of experiments on the annealing of radiation damage in copper have been performed, and two of these will be considered in some

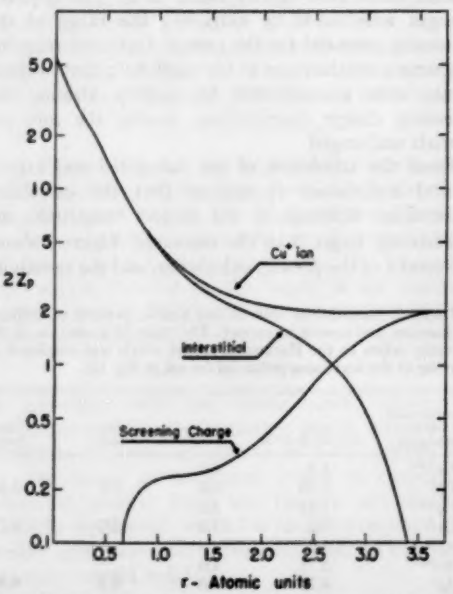


FIG. 2. Twice the effective nuclear charges as functions of radius for singly ionized copper, for the assumed screening charge distribution, and for the resulting interstitial potential.

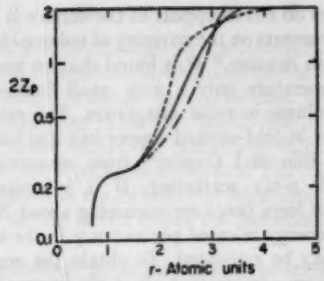


FIG. 3. Twice the effective nuclear charges for three screening charge distributions as functions of radius. Solid curve is the same as the corresponding curve of Fig. 2.

detail here. In addition to this work, the results of investigations on self-diffusion in the noble metals and on the quenching-in of vacancies in gold also yield valuable information on the properties of lattice defects in copper.

Simultaneous resistivity and calorimetric measurements on irradiated copper were carried out by Overhauser,[17,30] who studied in this manner the annealing spectrum in the temperature range extending from about −180°C to room temperature for samples irradiated at liquid nitrogen temperature. Some annealing was observed over the entire temperature range from −180°C to −40°C, but no unique activation energy could be assigned in that range. An initial bending of the resistivity *versus* bombardment curve was apparent, and this was attributed, in part, to thermal annealing at the bombardment temperature. This anneal was ascribed to a recombination of close pairs, and an activation energy of approximately 0.2 ev was assigned to that process.

At about −30°C Overhauser found a sharp annealing peak with an activation energy of 0.68 ev. An analysis of the kinetics of the anneal indicated that it was governed by a process of bimolecular recombination, and from this it was inferred that the anneal corresponded to a recombination of vacancies and interstitials. The simultaneous measurements of resistivity and of stored energy released during anneal lead to the value of 1.7 (cal/gram) per μohm-cm for the stored energy to resistivity ratio in this material at −30°C as well as in the lower temperature range. If one now uses a specific value for the energy of formation of a Frenkel defect in copper, the resistivity due to one atomic percent of vacancy-interstitial pairs may be obtained. Overhauser used energies of formation of 4 ev for interstitials and 1 ev for vacancies; the latter value is probably somewhat low, 1.3 to 1.5 ev being perhaps more nearly correct.[31] Taking 5.3 ev as the energy of formation of a Frenkel defect, the stored energy from one atomic percent concentration of pairs is 19 cal/gram. Hence, the resistivity per atomic percent of Frenkel pairs is 19/1.7=11.2 μohm-cm. When the calculated resistivity due to one atomic percent of vacancies is subtracted from the above, one arrives at the estimate of Overhauser, namely a resistivity increase of approximately 10 μohm-cm per atomic percent of interstitials in copper. This value is evidently in substantial disagreement with the results of the previous section.

Before discussing the second experiment, some difficulties in the interpretation which was outlined above will be indicated. If the anneal at −30°C is due to recombination of vacancies and interstitials, then the activation energy of the process must be that associated with motion of the more mobile of the two imperfections. The most recent measurement of the activation energy for self-diffusion in copper yielded the value 2.04 ev.[32] Since it is fairly certain that self-diffusion in copper proceeds by a vacancy mechanism, 2.04 ev is the sum of the energies of formation and of motion of vacancies in copper. Subtracting the calculated energy of formation of vacancies, one arrives at an energy of about 0.7 ev for motion of vacancies in copper. Further evidence that the energy of motion of vacancies is approximately 0.7 ev is provided by the measurements of Kauffman and Koehler,[33] who found that in gold the activation energies of formation and motion of vacancies are 1.28 ev and 0.68 ev, respectively. Although similar measurements on copper have not been reported, the fact that the two metals have the same crystal structure, have almost identical activation energies for self-diffusion, and are otherwise similar in many respects, gives support to the belief that the activation energy for motion of vacancies in copper is about 0.7 ev. If, then, the energy of 0.68 ev of the −30°C anneal is associated with motion of vacancies, recombination of Frenkel pairs at that temperature can occur only if interstitials are less mobile than vacancies. However, Huntington[34] recently calculated the energy of motion of interstitials in copper and found that it is rather small, in the neighborhood of 0.1 or 0.2 ev. It then follows that interstitial-vacancy recombination should occur at a temperature considerably below −30°C, and it is the low-temperature anneal reported by Overhauser, to which he assigned a tentative activation energy of 0.2 ev, that might well be due to recombination.

A more recent experiment[35] throws considerable light on these questions. Copper, silver, and gold were bombarded with 12-Mev deuterons at liquid helium temperature, and the resistivity increases were determined as functions of the integrated beam intensity. Following the irradiation the samples were permitted

[30] A. W. Overhauser, Phys. Rev. 90, 393 (1953).
[31] H. B. Huntington, Phys. Rev. 61, 325 (1942).
[32] Kuper, Letaw, Slifkin, Sonder, and Tomizuka, Phys. Rev. 96, 1224 (1954).
[33] J. W. Kauffman and J. S. Koehler, Phys. Rev. 97, 555 (1955).
[34] H. B. Huntington, Phys. Rev. 91, 1092 (1953).
[35] Cooper, Koehler, and Marx, Phys. Rev. 97, 599 (1955).

to warm to room temperature, and the resistivity annealing spectrum was obtained.

A small amount of curvature of the resistivity *versus* bombardment curve was observed even at these low temperatures, the effect being attributed to so-called radiation annealing. The annealing spectrum of copper observed by Cooper, Koehler, and Marx exhibited a very marked annealing peak at about 40°K, at which temperature about half of the initial resistivity increase annealed out. Additional resistivity decrease was observed in the range 45°K to 220°K, which was followed by another peak corresponding to the one found by Overhauser.

The estimated activation energy for the 40°K anneal is about 0.1 ev, and this is approximately the same as the activation energy for motion of interstitials as calculated by Huntington. It is, therefore, reasonable to associate this low-temperature anneal with motion of interstitials. Since the activation energy for motion of vacancies in copper is very nearly identical to the energy of 0.68 ev assigned to the −30°C anneal observed in both experiments, it is concluded that this anneal arises from the motion of vacancies.

According to this interpretation, however, there should be essentially no interstitials in the lattice at the higher temperature, and the bimolecular nature of the process requires an explanation other than vacancy-interstitial recombination. It has been suggested[35] that as vacancies become mobile they will combine to form divacancies, and that this process accounts for the second-order kinetics. Calculations of Bartlett and Dienes[36] indicate that such combined pairs should be stable and also considerably more mobile than single vacancies. Because of their high mobility, the effect of divacancies on the resistivity will not be subject to measurement, for these double vacancies will anneal out practically as soon as they are formed, moving to traps such as dislocations or grain boundaries, or giving rise to small holes by clustering.

Further confirmation of a divacancy process comes from the work of Kauffman and Koehler on the annealing of vacancies in gold. The annealing here also proceeds by a bimolecular reaction,[37] and since vacancies only are present, formation of divacancies is the only consistent interpretation of these findings.

Let us now see if the above interpretation of the annealing spectrum of irradiated copper leads to agreement with the calculated resistivities due to vacancies and interstitials. Bartlett and Dienes obtained for the energy of dissociation of a divacancy 0.6 ev, and this is then the energy released when two vacancies combine to form such a pair. Additional energy will, however, be released as the divacancies anneal out at −30°C either by going to trapping sites or by the formation of small pores. That most of the divacancies do not disappear at the surface is indicated by measurements on the recovery of volume change due to radiation damage.[38] It is found that on warming to room temperature only a very small fraction of the original volume increase disappears. The existence of small holes in cold-worked copper has also been postulated by Blin and Guinier[39] from measurements of small-angle x-ray scattering. If it is assumed that divacancies form pores encompassing about 200-lattice sites, the energy released per vacancy in the annealing process may be estimated. To obtain the energy of a small hole we assume that all of the energy is due to the surface,[40] and taking as the surface tension that of the molten metal, the surface energy of such a hole is then about 10^{-11} erg. The energy released per vacancy as they combine to form divacancies which, in turn, cluster into these holes, is then about 0.9 ev, assuming that the energy of formation of a vacancy is 1.3 ev. The measured stored energy to resistivity ratio of 1.7 (cal/gram) per μohm-cm now leads to a resistivity increase due to one atomic percent of vacancies of about 1.6 μohm-cm, in fair agreement with the calculated values.

A rough estimate of the resistivity due to interstitials may also be obtained. Let us assume that, immediately following bombardment at helium temperature, interstitials and vacancies exist in equal concentrations. It follows that annealing of interstitials with a simultaneous decrease of residual resistivity by half implies that the resistivity due to these imperfections is probably of the same order of magnitude and definitely less than that due to vacancies, since interstitials will combine at least in part with vacancies. This conclusion is also in qualitative agreement with the present results.

Although so far the foregoing interpretation of the annealing spectrum appears to be internally consistent and in agreement with a variety of theoretical and experimental results, there remain several difficulties which are unresolved. Overhauser reported that the ratio of energy released to resistivity change was the same in the region −100°C to −40°C as at −30°C, an observation which would indicate that the same process is responsible for the annealing over the entire temperature range studied by him. This conclusion may be difficult to reconcile with a process other than vacancy-interstitial recombination.

Another discrepancy arises when one calculates the expected resistivity increase due to bombardment with that which is actually observed. Multiplication of the estimated resistivity due to Frenkel defects by the number of defects as calculated by Harrison and Seitz[4] yields values about three times as large as those measured by Cooper, Koehler, and Marx.

[35] J. H. Bartlett and G. J. Dienes, Phys. Rev. 89, 848 (1953).
[37] J. W. Kauffman and J. S. Koehler (private communication).
[38] H. A. Kierstadt, Phys. Rev. 98, 245(A) (1955).
[39] J. Blin and A. Guinier, Compt. rend. 236, 2150 (1953).
[40] S. Machlup, Phys. Rev. 98, 1556(A) (1955).

VI. CONCLUSIONS

The results of calculations on the resistivity increase due to substitutional impurities and due to vacancies and interstitials in copper were presented. The partial wave method was used to obtain the differential scattering cross section for scattering of conduction electrons by these point imperfections. The calculations were carried out in the free-electron approximation, and relaxation of the lattice about the imperfections was neglected.

The scattering potential was approximated by taking the difference of the Hartree potentials with and without the imperfection present, and suitably screening this potential by an appropriate charge distribution. The screening charge distribution was adjusted until the calculated phase shifts satisfied the Friedel sum criterion, Eq. (5). Approximate agreement with resistivity changes measured by Linde was obtained under these assumptions, although the calculated resistivities are consistently too high. The most satisfactory screening charge distributions about impurities with valences greater than copper are very similar to the charge distributions corresponding to bound states in the free ion. As anticipated by Friedel, the details of the scattering potential do not influence the results very much as long as the phase shifts satisfy the Friedel sum condition.

The resistivity due to interstitials in copper was determined in the same manner. The calculated resistivity, 1.4 μohm-cm per atomic percent of interstitials, is practically the same as the calculated resistivity due to an equal concentration of vacancies. In view of the results in the case of substitutional alloys, it is believed, however, that the true resistivity increase due to Frenkel defects is overestimated by these calculations, and that a reasonable value lies between 1.5 and 2.0 μohm-cm per atomic percent of vacancy-interstitial pairs.

This result is in substantial disagreement with an earlier estimate of Overhauser. An attempt is made to reinterpret Overhauser's results in the light of more recent work of Cooper, Koehler, and Marx. It is suggested that at 40°K interstitials become mobile, and that the annealing peak observed by Overhauser and also by Cooper, Koehler, and Marx at about -30°C is due to motion of vacancies. These presumably combine to form divacancies, thereby accounting for the bimolecular nature of the anneal, and the divacancies, in turn, rapidly combine to form larger clusters. With this interpretation the resistivity due to vacancies as deduced from Overhauser's measurements is in fair agreement with the present estimates. This interpretation of the annealing spectrum is also consistent with the estimated resistivity due to interstitials.

Furthermore, the approximate mobility energy of interstitials as obtained from the 40°K anneal is in agreement with theoretical estimates of Huntington. The energy of motion of 0.68 ev associated with the -30°C anneal is in very close agreement with mobility energies of vacancies as deduced from self-diffusion measurements and quenching experiments on gold. There remain, however, several difficulties which have not been resolved.

VII. ACKNOWLEDGMENTS

It is a pleasure to express my gratitude to Mr. W. A. Harrison, who took an active interest in certain aspects of this problem, for his valuable help. Dr. R. Rubenstein, who had devoted much time and effort to the writing of a program by which the ILLIAC could solve the radial wave equation, very gratiously made this program available for use in the present investigation; the calculations thus became practicable. In this connection much credit is due to Mrs. M. C. Huse, who made expert use of the program. I am indebted to Professor J. S. Koehler and Professor F. Seitz, who suggested the problem, for their continuing interest and numerous enlightening discussions. I have also profited from several discussions with Dr. S. Machlup and Dr. Y. Yafet.

Note added in proof.—Recently, Jongenburger [Nature **175**, 545 (1955) hereafter referred to as J] reported results of a calculation of the change in resistivity due to interstitials in copper, using a square-well potential as an approximation to the scattering potential of the interstitial. He found that $\Delta\rho_i$, the resistivity increase due to the interstitial, is between 1 and 2 μohm-cm/atomic percent, in substantial agreement with the results reported here. The contribution to $\Delta\rho$ arising from the displaced nearest and next nearest neighbors was also estimated with the result that $\Delta\rho_d$, the resistivity increase due to these displacements, is 3.5 μohm-cm/atomic percent, leading to a total resistivity due to interstitials of about 5 μohm-cm/atomic percent. The writer believes, however, that, for the following reasons, assumption 4 is valid and that the influence of the displaced atoms was considerably overestimated by Jongenburger.

(1) It is assumed in J that an ion which has suffered a *static* mean-square displacement from its equilibrium position contributes just as much to the resistivity as an ion in thermal vibration at a high temperature ($T>\theta$) with the same mean-square displacement. In the latter case, however, primarily phonons of maximum wave vector are excited, so that in each collision an electron suffers a large momentum change. In the case of a static displacement the same is not true and a much smaller average momentum transfer is to be expected.

(2) In J the mean-square displacements of all the nearest and next nearest neighbors to the interstitial are added to give a total mean-square displacement about the imperfection. In this approximation all interference effects between the several displaced atoms are neglected, an approximation which is probably not

very good when the displaced atoms are nearest neighbors to each other. Including interference terms need not, of course, necessarily reduce the scattering, and may, when treated in detail, even lead to an enhancement of $\Delta\rho_D$.

(3) The results of J would indicate that $|m|^2$ and $|M|^2$, in the notation of Dexter,[16] are, in arbitrary units, of order 1.5 and 3.5, respectively. Now it was shown in reference 16 that the matrix elements m and M are, in the case of interstitials, very likely of opposite sign. Since the resistivity is in fact related to $|m+M|^2$, the above values for $|m|^2$ and $|M|^2$ may well lead to a rather small resistivity change due to interstitials, perhaps of order 0.5 μohm-cm/atomic percent. This crude argument is used here only to point out that it is essential to include interference effects when scattering from the imperfection and from the associated lattice distortion appears to be of the same magnitude when considered independently. [See also D. L. Dexter, Phys. Rev. **98**, 543 (1955).]

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

de Haas-van Alphen Effect in Arsenic

TED G. BERLINCOURT
United States Naval Research Laboratory, Washington, D. C.
(Received March 24, 1955)

The de Haas-van Alphen effect has been studied in single crystals of rhombohedral (metallic) arsenic in the liquid helium temperature range and in magnetic fields up to 25 kilogauss. The effect in arsenic is characterized by long-period (up to $\sim 4\times10^{-4}$ gauss^{-1}) oscillations upon which are superposed three short period (up to $\sim 6\times10^{-7}$ gauss^{-1}) terms. Analysis in terms of existing theory attributes the long-period oscillations to electrons occupying a constant energy surface in momentum or wave-number space which to a first approximation is an ellipsoid of revolution with a degeneracy energy $E_{el}=1.59\times10^{-14}$ erg. The pertinent electrons in the case of the short-period oscillations can be attributed to three identical ellipsoidal constant-energy surfaces oriented so as to satisfy the trigonal symmetry of the arsenic lattice and having a degeneracy energy $E_{eh}=29.4\times10^{-14}$ erg. In addition, the electronic effective masses have been evaluated. A method for growing arsenic single crystals is described.

INTRODUCTION AND THEORY

THE de Haas-van Alphen effect is characterized by a periodic dependence of the magnetic susceptibility of metal single crystals upon reciprocal magnetic field at low temperatures. Since its discovery by de Haas and van Alphen[1] in bismuth, the effect has been observed in fourteen other metal single crystals,[2] and correlations have been established between the susceptibility oscillations and similar oscillations in the magnetoresistance,[3-5] Hall effect,[6,7] thermoelectric effect,[8] and thermal conductivity.[9]

The discovery of short-period susceptibility oscillations in arsenic was reported briefly in an earlier communication[10] by the author. In addition, much-longer-period oscillations have subsequently been detected, and the present paper concerns detailed investigations of both the long- and short-period terms.

The theory of the de Haas-van Alphen effect was developed by Peierls,[11] Blackman,[12] and Landau[13] and is based upon a free-electron model. In the presence of a magnetic field the electronic energy levels coalesce into quantized levels of field-dependent spacing and degeneracy. As the field is varied, the electrons are redistributed among these levels in such a way as to give rise to the susceptibility oscillations. The effect of the periodic electric field of the lattice is taken into account by the introduction of anistropic effective electronic masses, and in many cases (including arsenic) the pertinent constant-energy surfaces in momentum or wave-number space may be approximated by one or more ellipsoids. Such approximately ellipsoidal pockets of electrons (or holes) presumably exist where the Fermi surface overlaps (or underlaps) Brillouin zone boundaries. The resulting high curvature and correspondingly low degeneracy energy and low effective electronic masses favor easy detection of the de Haas-van Alphen effect.

Equation (1), obtained from Landau's theory (see

[1] W. J. de Haas and P. M. van Alphen, Leiden Comm. No. 212 A (1930).
[2] D. Shoenberg, Trans. Roy. Soc. (London) **245**, 1 (1952). This reference discusses most of the work before 1952.
[3] P. B. Alers and R. T. Webber, Phys. Rev. **91**, 1060 (1953).
[4] Ted G. Berlincourt, Phys. Rev. **91**, 1277 (1953).
[5] T. G. Berlincourt and J. K. Logan, Phys. Rev. **93**, 348 (1954).
[6] Laird C. Brodie, Phys. Rev. **93**, 935 (1954).
[7] Reynolds, Leinhardt, and Hemstreet, Phys. Rev. **93**, 247 (1954).
[8] M. C. Steele and J. Babiskin, Phys. Rev. **94**, 1394 (1954).
[9] J. Babiskin and M. C. Steele, Phys. Rev. **96**, 822 (1954). Also, P. B. Alers, Phys. Rev. **98**, 1180(A) (1955).
[10] Ted G. Berlincourt, Phys. Rev. **92**, 1068 (1953).
[11] R. Peierls, Z. Physik **81**, 186 (1933).
[12] M. Blackman, Proc. Roy. Soc. (London) **A166**, 1 (1938).
[13] L. D. Landau, see Appendix of D. Shoenberg, Proc. Roy. Soc. (London) **A170**, 341 (1939).

reference 2), gives the difference between the magnetic susceptipilities ($\Delta\chi$) per unit mass in two directions at right angles to each other for the case of ellipsoidal constant energy surfaces. (Recent modifications of this theory and methods for evaluating electronic parameters from the experimental data will be discussed later.)

$$\Delta\chi=\sum\frac{A\Delta m}{\rho}\left\{\frac{\pi^2}{6}\left(\frac{k}{E_0}\right)^{\frac{1}{2}}-\frac{1}{T^{\frac{1}{2}}}\left(\frac{2\pi^2kT}{\beta H}\right)^{\frac{1}{2}}\times\sum_{p=1}^{\infty}\frac{(-1)^{p+1}\sin(2\pi pE_0/\beta H-\pi/4)}{2p^{\frac{1}{2}}\sinh(2\pi^2pkT/\beta H)}\right\},\quad(1)$$

provided $E_0\gg kT$ and $E_0\gg\beta H$. As will be evident later, the latter is not fulfilled for the long period oscillations at the highest fields of this investigation. A is a constant given by

$$A\equiv\frac{e^2E_0}{\pi^4c^2\hbar(2k)^{\frac{1}{2}}(m')^{\frac{1}{2}}},\quad(2)$$

ρ is the density, β is a double effective Bohr magneton given by

$$\beta=e\hbar/m''c,\quad(3)$$

and m', m'', and Δm are functions of the relevant effective masses. These functions depend upon the geometry of the experiment and the shape of the electronic constant energy surface in momentum space for the crystal under investigation. E_0 is the energy at this surface measured from the bottom of the relevant zone in the case of electrons (or from the top in the case of holes) and is given by a quadratic function of the momenta. The first summation sign in Eq. (1) is included in order to account for the existence of several ellipsoidal surfaces characterized by different parameters and giving rise to beats. As will be borne out by the experimental data, it is possible to account almost completely for the observed oscillatory terms in arsenic by the introduction of four ellipsoids. At this point it is convenient to write the pertinent equations for these ellipsoids, postponing a discussion of their applicability to a later section.

1. Long-Period Oscillations

The long-period oscillations in arsenic will be interpreted in terms of a single ellipsoid of the form

$$E_{0l}=\frac{p_x^2+p_y^2}{2m_\perp}+\frac{p_z^2}{2m_{\parallel}},\quad(4)$$

where p_x, p_y, and p_z are the components of momentum parallel respectively to the binary axis, the bisectrix axis, and the trigional axis. $m_\perp$ and $m_{\parallel}$ are the effective electronic masses perpendicular and parallel respectively to the trigonal axis, and the subscript l refers to the long-period term.

$$m_l'=m_\perp(m_{\parallel})^{\frac{1}{2}},\quad(5)$$

$$(\Delta m)_l=m_\perp-m_{\parallel},\quad(6)$$

$$\beta_l=\frac{e\hbar}{cm_\perp(m_{\parallel})^{\frac{1}{2}}}(m_\perp\sin^2\phi+m_{\parallel}\cos^2\phi)^{\frac{1}{2}},\quad(7)$$

where ϕ is the angle between the trigonal axis and the magnetic field.

2. Short-Period Oscillations

The three-ellipsoid scheme proposed by Shoenberg for bismuth and antimony is appropriate, and the pertinent equations are copied below with only slight changes in notation.

$$2m_0E_{0s}=\alpha_1p_x^2+\alpha_2p_y^2+\alpha_3p_z^2+2\alpha_4p_yp_z,\quad(8)$$

and in addition the two ellipsoids derived from this one by rotations of $\pm\frac{2}{3}\pi$ about the p_x axis. The subscript s refers to the short-period oscillations, m_0 is the free-electron mass, and

$$\alpha_1=\frac{m_0}{m_1},\quad\alpha_2=\frac{m_0m_3}{m_2m_3-m_4^2},\quad\alpha_3=\frac{m_0m_2}{m_2m_3-m_4^2},$$

$$\alpha_4=\frac{m_0m_4}{m_2m_3-m_4^2}.$$

These ellipsoids lead to three periodic terms in Eq. (1) which, for the modes of measurement indicated, have the following values for β_s and $(\Delta m)_s$ provided H is in a horizontal plane and ϕ is the angle between the trigonal axis and the field.

(a) Binary axis vertical, trigonal axis horizontal

$$\beta_1=\lambda(m_3\cos^2\phi+m_2\sin^2\phi+2m_4\sin\phi\cos\phi)^{\frac{1}{2}},$$
$$\beta_2=\beta_3=\lambda(m_3\cos^2\phi+\tfrac{1}{4}(3m_1+m_2)\sin^2\phi-m_4\sin\phi\cos\phi)^{\frac{1}{2}},\quad(9)$$

$$(\Delta m)_1\sin\phi\cos\phi=\tfrac{1}{2}(m_2-m_3)\sin2\phi+m_4\cos2\phi,$$
$$(\Delta m)_2\sin\phi\cos\phi=(\Delta m)_3\sin\phi\cos\phi=\tfrac{1}{2}(\tfrac{1}{4}m_1+\tfrac{1}{4}m_2-m_3)\sin2\phi-\tfrac{1}{2}m_4\cos2\phi.\quad(10)$$

It should be noted that for this mode of measurement, ϕ and $\pi-\phi$ are not equivalent. As in Shoenberg's work, the sense of ϕ has been so chosen that ϕ increases from zero as the field turns away from the trigonal axis (0001) into the acute-angled sector between (0001) and $(10\bar{1}1)$.

(b) Binary and trigonal axes horizontal

$$\beta_1=\lambda(m_3\cos^2\phi+m_1\sin^2\phi)^{\frac{1}{2}}$$
$$\beta_2,\beta_3=\lambda[m_3\cos^2\phi+\tfrac{1}{4}(m_1+3m_2)\sin^2\phi\pm\sqrt{3}m_4\sin\phi\cos\phi]^{\frac{1}{2}},\quad(11)$$

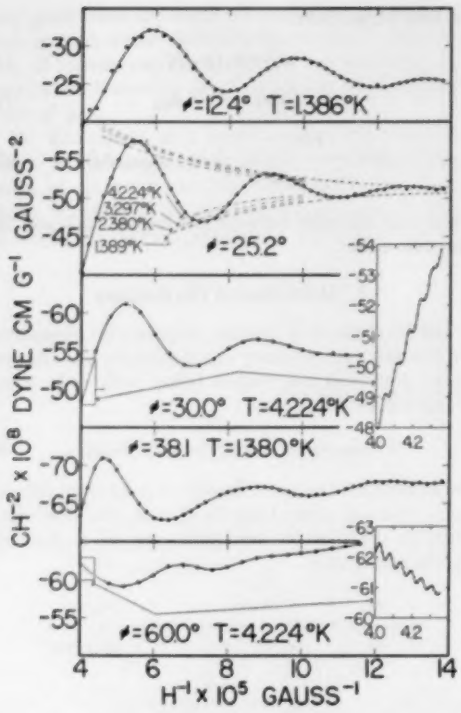


FIG. 1. As 1 (impure). CH^{-2}, which is proportional to the magnetic anisotropy $\Delta\chi$, *versus* H^{-1} for mode (*b*), binary and trigonal axes horizontal. ϕ is the angle between the trigonal axis and H. The short-period oscillations were present in general but are not plotted in detail in all cases.

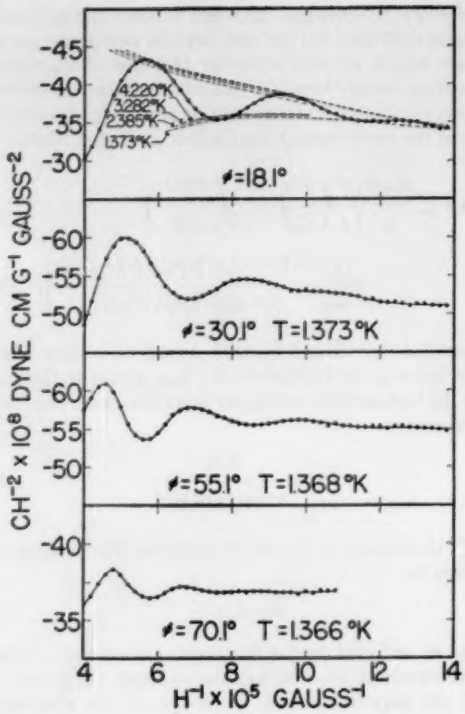


FIG. 2. As 1 (impure). CH^{-2} *versus* H^{-1} for mode (*a*), binary axis vertical, trigonal axis horizontal. Though not indicated, the short-period oscillations were present in general.

$$(\Delta m)_1 \sin\phi\cos\phi = \tfrac{1}{2}(m_1 - m_3)\sin 2\phi,$$

$$(\Delta m)_2 \sin\phi\cos\phi,\ (\Delta m)_3 \sin\phi\cos\phi = \tfrac{1}{2}(\tfrac{1}{4}m_1 + \tfrac{3}{4}m_2 - m_3)\sin 2\phi \pm \tfrac{1}{2}\sqrt{3}m_4\cos 2\phi. \quad (12)$$

In both (*a*) and (*b*) above,

$$\lambda = \frac{e\hbar}{c[m_1(m_2m_3 - m_4^2)]^{\frac{1}{3}}}, \quad (13)$$

and the m' of Eq. (2) is given by

$$m' = [m_1(m_2m_3 - m_4^2)]^{\frac{1}{3}}. \quad (14)$$

EXPERIMENTAL WORK

1. Torsion Balance Method

The torsion balance method (which is described in detail elsewhere[9,14]) was used to measure the couple C per unit mass in a vertical direction acting on the crystal in a homogeneous magnetic field. This couple is related to $\Delta\chi$ [as given by Eq. (1)] by the expression:

$$C = \Delta\chi H^2 \sin\phi\cos\phi, \quad (15)$$

where H is the magnetic field strength and ϕ is the angle measured in a horizontal plane between a given crystallographic axis and the magnetic field. Three modes of measurement were employed in this investigation, (*a*) binary axis vertical, trigonal axis horizontal, (*b*) binary and trigonal axes horizontal, (*c*) trigonal axis vertical.

Angular displacements of the crystal never exceeded 1° and displacements of approximately 0.0005° were detectable. Values of ϕ were corrected for these displacements, and measurements of the short-period terms were carried out over small ranges of field in order to avoid complications arising from their sensitive orientation dependence. Absolute values of CH^{-2} may be in error by as great as 2 percent because of possible errors in balance calibration, but relative values should be accurate to about 0.2 percent at high fields.

Temperatures were determined by observation of the vapor pressure of the liquid helium bath with mercury and oil manometers used in conjunction with the Mond vapor pressure tables.[15]

2. Magnet

An Arthur D. Little electromagnet capable of rotation about a vertical axis supplied horizontal magnetic fields up to 25 kilogauss in strength. Pole pieces 5.75

[14] Ted G. Berlincourt, Phys. Rev. **88**, 242 (1952).

[15] H. van Dijk and D. Shoenberg, Nature **164**, 151 (1949).

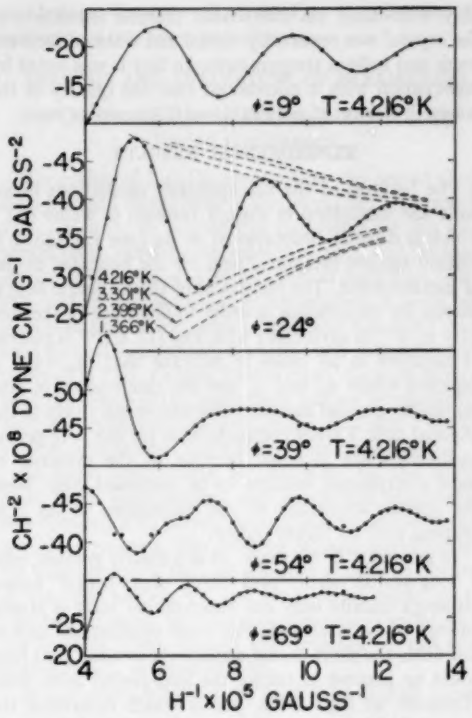


FIG. 3. As 2 (pure). CH^{-2} *versus* H^{-1} for mode (*b*), binary and trigonal axes horizontal. Note the beats for $\phi=39°$ and $\phi=54°$. The short-period oscillations were present in general.

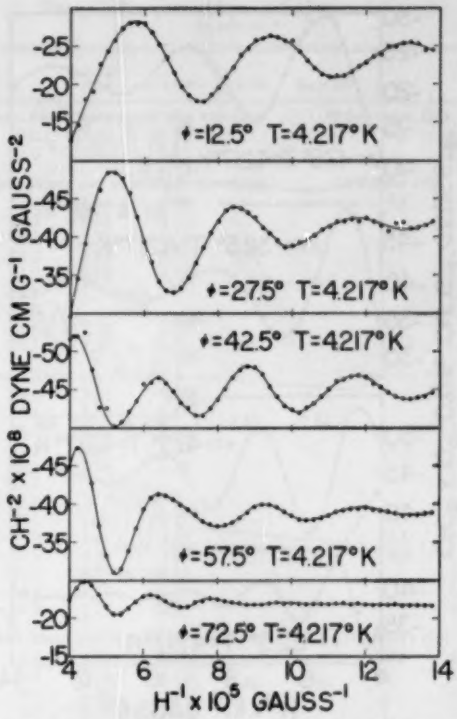


FIG. 4. As 2 (pure). CH^{-2} *versus* H^{-1} for mode (*a*), binary axis vertical, trigonal axis horizontal. Note the beats for $\phi=42.5°$ and $\phi=57.5°$. The short-period oscillations were present in general.

inches in diameter and a gap of 1.75 inches assured field homogeneity to within a few gauss over the volume of the crystal. The magnet was calibrated with a nuclear fluxmeter to an accuracy of better than 0.02 percent, and time variations of the field were limited to less than 5 gauss during the course of a single measurement.

3. Arsenic Crystals

Two arsenic crystals were used in this investigation. The first crystal, As 1, (weighing 13.7 milligrams) was a well-formed hexagonal plate cleaved at a small connecting link from a large crystalline mass of Fisher A892 "Purified Arsenic." A spectrochemical analysis kindly carried out by S. Cress of this laboratory revealed the following impurities: 0.1 to 1 percent Sb, 0.01 to 0.1 percent Bi, 0.01 to 0.1 percent Si, 0.001 to 0.01 percent, Cu, 0.001 to 0.01 percent Pb, and traces of Mg. The second crystal, As 2, (weighing 14.0 milligrams) was grown from the vapor with Johnson Matthey and Company, JM 640, Laboratory No. 6126 very high-purity arsenic. The spectroscopic report listed faintly visible lines from Ag, Cu, Mg, and Si.

Difficulties were anticipated in growing a crystal because of the many allotropic modifications of arsenic and the fact that it melts only at relatively high pressures and temperatures (see Mellor[16]). However, since it sublimes readily at moderate temperatures, growth from the vapor is feasible.[17] The method finally adopted will be described in some detail.

A simple furnace was constructed in the following way: A slot $\frac{1}{32}$ inch wide, 5 inches long and open at one end was cut in a $\frac{3}{4}$-inch thick slab of transite, and along both sides of the slot small holes were drilled 3 mm apart and 1.5 mm from the edge. Nichrome heater wire was threaded through these holes in spiral fashion. This piece was then placed on a 1-inch thick transite slab, and two small $\frac{3}{4}$-inch thick transite blocks were placed on top, each covering nearly half the length of the slot leaving a gap of about $\frac{1}{4}$ inch at the center. The nichrome wires were spread apart slightly at this gap to provide a "cool spot." Roughly 20 milligrams of arsenic were sealed under high vacuum into a 7-mm o.d. quartz tube about 3 inches long, which was in turn placed into the furnace slot. Heat was supplied until all of the arsenic was vaporized. A crystal was then nucleated by touching the quartz tube at the "cool spot" with a

[16] J. W. Mellor, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry* (Longmans, Green and Company, London, 1929), Vol. 9, p. 16.

[17] McLennan, Niven, and Wilhelm, Phil. Mag. 6, 666 (1928).

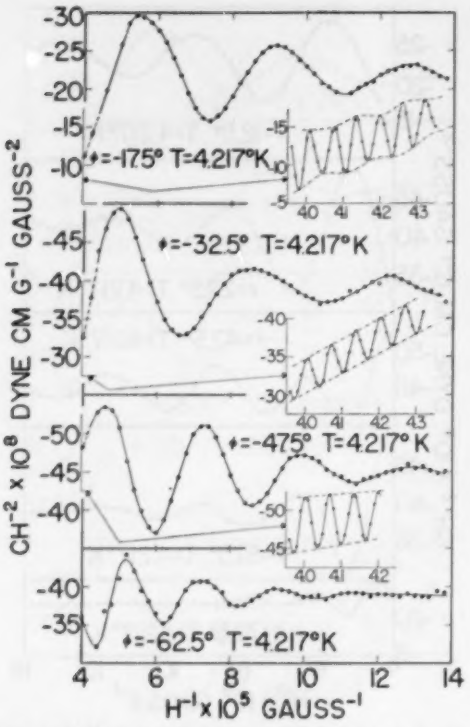


FIG. 5. As 2 (pure). CH^{-2} *versus* H^{-1} for mode (*a*), binary axis vertical, trigonal axis horizontal. Note the much larger amplitude for the short-period terms than was observed for As 1.

cool, sharp metal probe for a few seconds, and decreasing the power supplied to the furnace. The growth of the crystal was observed visually, the power being adjusted to obtain a suitable growth rate. In order to condense only the rhombohedral form of arsenic, it appeared to be necessary to vaporize all of the arsenic and then condense it at a relatively high temperature. Attempts to sublime a small part of a large quantity of arsenic from one region to a cooler region produced unwanted allotropic forms. It should be mentioned that the temperatures used often exceeded the softening point of Pyrex, which was used in place of quartz in early attempts.

As 2, produced in the above manner, displayed well-formed facets, and a goniometric investigated yielded reflections from the [0001], [01$\bar{1}$2], [10$\bar{1}$1], [20$\bar{2}$1], [10$\bar{1}$4] planes. All except the third of these forms were noted by Palache[18] in a natural arsenic crystal. Despite the fact that As 2 was the best crystal obtained, a Laue back-reflection x-ray picture taken after the magnetic measurements had been completed revealed considerable asterism. Since the crystal was attached to the torsion balance with Duco cement, strains were probably introduced via differential thermal expansion as the crystal was repeatedly cooled and warmed between room and helium temperatures. In fact it was noted by observation with a microscope that the quality of the facets deteriorated noticeably with successive runs.

[18] Charles Palache, Am. Mineralogist 26, 709 (1941).

EXPERIMENTAL RESULTS

The features of the susceptibility oscillations in arsenic are illustrated in Figs. 1 through 6, where CH^{-2} which is directly proportional to $\Delta\chi$ [see Eq. (15)], is plotted against reciprocal field for the indicated modes of measurement. The reciprocal of the magnetic field is chosen for the abscissa in order to illustrate the periodicity in H^{-1} in agreement with Eq. (1). CH^{-2} is plotted as negative in all cases to indicate that $(\chi_{\parallel}-\chi_{\perp})$ is negative where $\chi_{\parallel}$ and $\chi_{\perp}$ are the components of susceptibility parallel and perpendicular respectively to the trigonal axis. Rather complete data for the long-period oscillations are included because of the presence of some unexplained features to be discussed later. Note the greater amplitude of the oscillations in As 2 in keeping with its higher purity.

In general, CH^{-2} is made up of a steady portion, plus a long-period term, plus three short-period terms, although usually only one short-period term is strong enough to be measured. For some orientations and at low fields the short-period terms were so weak that they could be ignored in taking the long-period term data. However, at high fields, points which determine the long-period curves were taken as the mean of the short-period oscillations over small intervals of field. It should be remembered that Figs. 1 through 5 are in most parts schematic in that the short-period terms were usually present. They were not traced out in detail over the entire range of fields as this procedure would have been prohibitively time-consuming. Note the presence of beats in the blow-up of the short-period

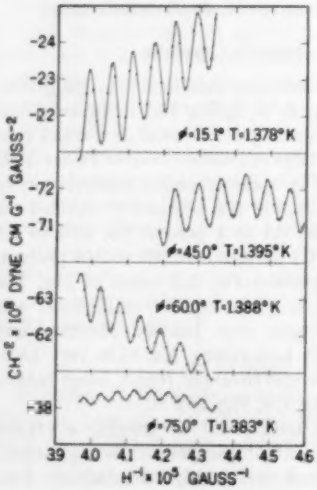


Fig. 6. As 1 (impure). CH^{-2} *versus* H^{-1} for mode (*b*), binary and trigonal axes horizontal. The shift of the mean of the short-period term arises from the long-period term illustrated in Figs. 1 through 5.

oscillations for $\phi=-17.5°$ in Fig. 5 indicating the existence of two short-period terms.

ANALYSIS AND EVALUATION OF PARAMETERS

1. The One Ellipsoid Scheme and the Long-Period Oscillations

Inspection of Eq. (1) reveals that the period of the susceptibility oscillations in reciprocal field is given by β/E_0. In order to deduce this parameter for the long-period term, envelope curves were drawn, and the mean of the oscillations was ascertained in each case. Then values of reciprocal field at which maxima, minima, and crossover points (or zeros of the sine function) occur were plotted against odd, even, and half-integers respectively as in Figs. 7 and 8. Values for β_1/E_{01} were deduced from the slopes of the linear plots and are listed together with the corresponding amplitudes in Table I for the various orientations and modes of measurement.

At some orientations, as illustrated by the dashed lines, beats were present, and insufficient information was obtainable to permit determinations of the periods of the components. The presence of beats could indicate the inadequacy of the one ellipsoid scheme as outlined in the first section for the long period term, or the existence of small inclusions of different orientation in the crystals (even though not detected by the x-ray examination), or possibly both.

In any event, a test of the fit of the available experimental data to the one ellipsoid scheme was accomplished by plotting the square of the period against $\cos^2\phi$ for the (a) and (b) modes of measurement. If Eq. (7) is applicable, a straight line should be obtained with points for both modes falling on the same line. Such plots appear for the two crystals in Figs. 9 and 10. Evidently the one-ellipsoid scheme is a fair approximation. Note from Figs. 9 and 10 that the impure crystal, As 1, exhibited periods noticeably different from the pure crystal and far exceeding any difference attributable to the possible orientation errors of ±2°.

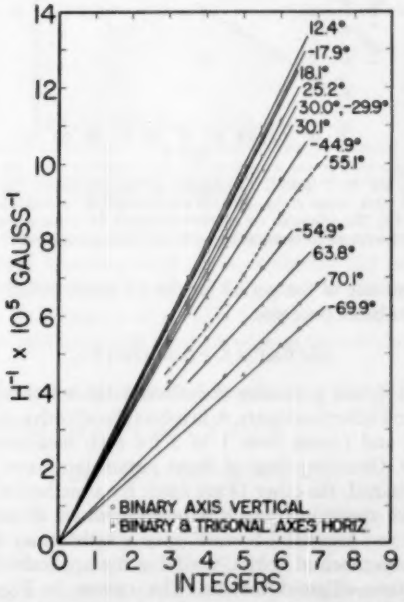


FIG. 7. As 1 (impure). Values of H^{-1} at which maxima, minima, and crossover points occur for the long period oscillations plotted against odd, even, and half integers respectively. Periods are calculated from the slopes of these lines and the phase is determined from the intercept on the integer axis.

TABLE I. Summary of data on the long-period term in arsenic.[a]

ϕ	$(\beta_1/E_{01})\times10^{5}$	$a_1\times10^{9}$	$\beta_1\times10^{19}$	$E_{01}\times10^{24}$	x_1	T
As 1, binary and trigonal axes horizontal					20 September 1954	
30.0	3.59	1.85			7.3	4.224
60.0	2.24	0.55				4.224
As 1, binary and trigonal axes horizontal					20 October 1954	
12.4	4.00	3.04			5.9	1.386
25.2	3.73	2.16				4.224
25.2	3.73	2.75	5.28	1.42		3.297
25.2	3.73	3.05				2.380
25.2	3.73	3.27			7.8	1.389
38.1	beats					1.380
51.0	beats					1.377
63.8	2.10	0.58				1.373
76.7	1.31	(0.1)				1.368
As 1, binary axis vertical					2 November 1954	
18.1	3.86	1.80				4.220
18.1	3.86	1.97	5.57	1.44		3.282
18.1	3.86	2.24				2.385
18.1	3.86	2.38			8.2	1.373
30.1	3.51	1.20				1.373
45.1	beats					1.372
55.1	beats					1.368
70.1	1.89	(0.2)				1.366
−17.9	3.89	2.58			7.1	1.363
−29.9	3.59	1.93			8.9	1.349
−44.9	2.99	1.63				1.357
−54.9	2.47	0.88				1.356
−69.9	1.73	(0.2)				1.342
As 2, binary and trigonal axes horizontal					17 January 1955	
9.0	3.74	4.18			3.0	4.216
9.0	3.74	4.78	5.90	1.58		3.301
9.0	3.74	5.42				2.403
9.0	3.74	5.92				1.380
24.0	3.50	6.86			4.5	4.216
24.0	3.50	7.98	5.64	1.59+		3.301
24.0	3.50	9.10				2.395
24.0	3.50	10.15				1.366
39.0	beats					4.216
54.0	2.48	(3.5)				4.216
69.0	1.86	1.00			2.2	4.216
As 2, binary axis vertical					24 January 1955	
12.5	3.68	4.35			3.2	4.217
27.5	3.30	4.45				4.217
42.5	beats					4.217
57.5	beats					4.217
72.5	1.80	0.47			3.6	4.217
−17.5	3.60	5.17			3.9	4.217
−32.5	3.30	3.87			5.3	4.217
−47.5	2.70	5.14			3.6	4.217
−62.5	2.14	1.42			4.5	4.217

[a] ϕ is the angle in degrees between the trigonal axis and the magnetic field; β_1/E_{01} is in gauss^{-1}; a_1 is the amplitude of the oscillations of CH^{-1} in dyne cm g^{-1} gauss^{-2} for $H^{-1}=8\times10^{-5}$; β_1 is in erg gauss^{-1}; E_{01} is in erg; x_1 is in °K; T is in °K; less accurate values are indicated by parentheses.

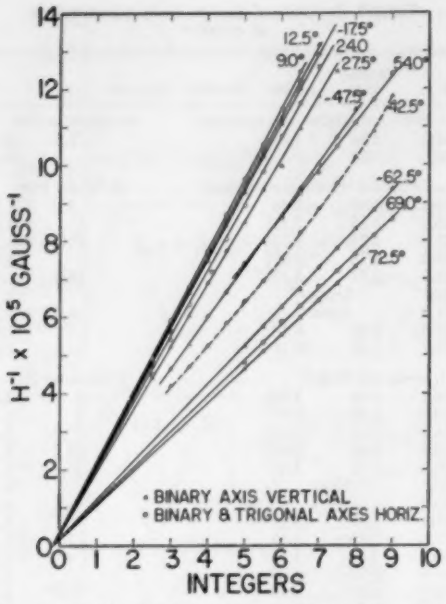


FIG. 8. As 2 (pure). Values of H^{-1} at which maxima, minima, and crossover points occur for the long-period oscillations plotted against odd, even, and half integers, respectively.

Additional support for the one-ellipsoid approximation, which predicts isotropy in the plane perpendicular to the trigonal axis, was obtained from measurements performed on both crystals with the trigonal axis vertical. These measurements yielded only very weak oscillations, too weak to analyze with any certainty, although the evidence suggested terms having periods of approximately 0.85×10^{-5} gauss^{-1} in As 2.

The phase of the long-period oscillations is also of interest and can be obtained from the intercepts on the integer axis of those plots in Figs. 7 and 8 for which an extrapolation to infinite field appears justified. For As 1, the observed phase $\delta=0.50\pi$ and for As 2, $\delta=0.31\pi$ whereas the theoretical phase as given in Eq. (1) is $\delta=-\pi/4$.

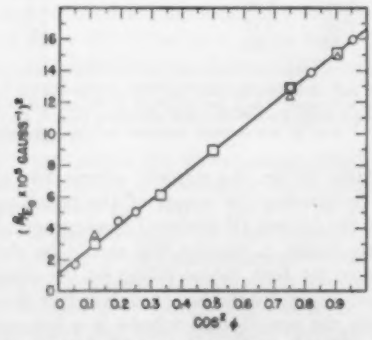


FIG. 9. As 1 (impure). The square of the period for the long-period term *versus* the square of the cosine of the angle ϕ between the trigonal axis and H. The circles correspond to measurement mode (b), the triangles correspond to mode (a) with ϕ positive, and the squares correspond to mode (a) with ϕ negative.

2. The Three-Ellipsoid Scheme and the Short-Period Oscillations

In determining periods for the short-period terms, a process similar to that described in the foregoing was adopted except that only crossover points were plotted against integers. Wherever beats occurred, the dominant period was taken as the mean period over several beat periods. The subordinate period could then be determined from the beat period. The values so obtained as well as the amplitude of the oscillations are tabulated in Table II.

In order to test the applicability of the bismuth-antimony type three-ellipsoid scheme to the short-period terms in arsenic, the square of the period was plotted against ϕ for measurement modes (a) and (b). Now according to Eqs. (9) and (11) the orientation dependence of the period for the ith short-period term should be of the form

$$(\beta_i/E_{0i})^2=f_i+g_i\sin(2\phi+\theta_i), \tag{16}$$

where f_i and g_i involve the effective masses as well as E_{0i} and other constants, θ_i involves the effective masses only, and i runs from 1 to 3 for each measurement mode. Once any four of these parameters have been determined, the other 14 are fixed. By a method of successive approximations a self-consistent set of parameters was ascertained which gave a satisfactory fit to the experimental points, verifying the applicability of the three-ellipsoid scheme. The curves in Figs. 11 through 14 were drawn by using these parameters, and it should be pointed out that very slight modifications (of the order of one percent) of the parameters produce noticeably poorer fits to the data. Note again that

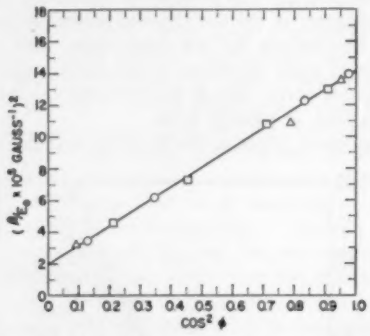


FIG. 10. As 2 (pure). The square of the period for the long-period term *versus* $\cos^2\phi$. The circles correspond to measurement mode (b), the triangles correspond to mode (a) with ϕ positive, and the squares correspond to mode (a) with ϕ negative.

TABLE II. Summary of data on the short period terms in arsenic.*

ϕ	$(\beta_s/E_{0s})\times10^{7}$	$a_s\times10^{9}$	$\beta_1\times10^{19}$	$E_{01}\times10^{14}$	x_1	T
As 1, binary axis vertical						13 August 1953
14.2	5.79	0.46				1.840
28.7	5.85	0.74				1.842
43.6	5.28	1.00				1.842
58.7	4.52	0.34				1.843
74.2	(3.9)	(0.05)				1.843
−2.7	5.44	0.38				4.224
−13.9	6.27	0.57				4.224
−28.4	6.73	0.56				4.224
−43.0	6.81	0.37				4.224
−58.3	6.27	0.72	(1.73)	(27.6)	(5.8)	4.224
−58.3	6.27	1.59			(6.6)	1.842
−73.7	5.34	0.21	(1.48)	(27.7)		4.224
−73.7	5.34	0.59				1.842
As 1, binary and trigonal axes horizontal						30 September 1954
15.1	5.99	1.06				1.378
30.0	6.42	0.241				4.224
30.0	6.42	0.332	1.86	29.0		3.307
30.0	6.42	0.427				2.380
30.0	6.42	0.556				1.389
45.0	6.41	0.754				1.395
60.0	5.72	0.181				4.224
60.0	5.72	0.256	1.69	29.6		3.301
60.0	5.72	0.350				2.380
60.0	5.72	0.470				1.388
75.0	4.77	0.110				1.383
−15.0	6.17	0.361				1.379
−22.4	6.32					1.381
−29.9	6.42	0.425				1.384
−44.9	6.37	1.16				1.390
−52.4	6.09					1.391
−59.9	5.68	1.32				1.392
−67.4	5.36					1.392
−75.0	4.69	0.280				1.393
As 2, binary and trigonal axes horizontal						17 January 1955
9.0	5.76	2.30				4.216
9.0	5.76	3.33	1.70	29.5		3.301
9.0	5.76	4.36				2.403
9.0	5.76	5.82				1.380
9.0	4.62	0.35				4.216
9.0	4.62	0.49				3.301

ϕ	$(\beta_s/E_{0s})\times10^{7}$	$a_s\times10^{9}$	$\beta_1\times10^{19}$	$E_{01}\times10^{14}$	x_1	T
As 2, binary and trigonal axes horizontal January 17, 1955—*Continued*						
9.0	4.62	0.69				2.403
9.0	4.62	0.91				1.380
24.0	6.11	1.82				4.216
24.0	6.11	2.69	1.79	29.3		3.301
24.0	6.11	3.42				2.395
24.0	6.11	4.45				1.366
24.0	5.02	0.57				4.216
24.0	5.02	0.82				3.301
24.0	5.02	1.13				2.395
24.0	5.02	(1.70)				1.366
39.0	6.35	1.53				4.216
39.0	4.50	0.30				4.216
54.0	5.95	3.40				4.216
69.0	5.25	1.91				4.216
84.0	4.16	(0.4)				4.216
−15.0	5.80	4.00				4.216
−30.0	6.30	1.38				4.216
−45.0	6.29	2.37				4.216
−60.0	5.61	2.90				4.216
−75.0	4.92	(1.3)				4.216
As 2, binary axis vertical						24 January 1955
7.5	5.52					4.217
12.5	5.53	(1.0)				4.217
18.0	3.81					4.217
35.0	5.60					4.217
42.5	5.60	1.43				4.217
50.0	5.37					4.217
57.5	4.85	(1.0)				4.217
65.0	4.23					4.217
72.5	3.80	0.10				4.217
80.0	3.47					4.217
−7.5	5.68					4.217
−17.5	6.21	3.17				4.217
−17.5	4.33	0.48				4.217
−25.0	6.59					4.217
−32.5	6.62	2.25				4.217
−47.5	6.57	3.27				4.217
−55.0	6.40					4.217
−70.0	5.57					4.217
−77.5	4.77	1.03				4.217
−85.0	4.36					4.217

* ϕ is the angle in degrees between the trigonal axis and the magnetic field; β_s/E_{0s} is in gauss^{-1}; a_s is the amplitude of the oscillations of CH^{-2} in dyne cm g^{-1} gauss^{-2} for $H^{-1}=4.1\times10^{-4}$; β_1 is in erg gauss^{-1}; x_1 is in °K; T is in °K; less accurate values are indicated by parentheses; wherever no entry appears under a_s, oscillations were counted.

small differences exist between the results for As 1 and As 2, the impure crystal again exhibiting the longer periods. No determination of the phases of the short-period oscillations was possible since extrapolation to infinite field could not be made with any certainty.

3. Effective Masses and Degeneracy Energies

Provided Eq. (1) is valid, a plot of

$$\log_e\{aT^{-1}[1-\exp(-4\pi^2kT/\beta H)]\}$$

(where a is the amplitude of the oscillations in CH^{-2} for a given term) against T for fixed H should yield a straight line of slope $-2\pi^2k/\beta H$. Thus, a first approximation to β may be obtained from the slope of a plot of $\log_e(aT^{-1})$ against T. This value of β may then be used in a plot of

$$\log_e\{aT^{-1}[1-\exp(-4\pi^2kT/\beta H)]\}$$

against T, yielding a second approximation to β. This process may be repeated until a value of β of the desired accuracy is obtained. Values of β so derived appear in Tables I and II, and examples (using $\log_{10}$) of such final plots appear in Fig. 15. The theoretical form of the temperature dependence of amplitude is verified by the linearity of these plots. [Incidentally, the harmonics appearing in Eq. (1) were not important at the fields and temperatures involved in these plots.]

Once β had been determined, E_0 was determined from the period β/E_0. Values for both long and short period terms in both crystals appear in Tables I and II.

Comparisons of Eq. (7) with Figs. 9 and 10, Eq. (9) with Figs. 11 and 12, and Eq. (11) with Figs. 13 and 14 permitted evaluation of the pertinent electronic effective masses. These are listed together with the best values for E_0, Table III. (The values for m_{11}/m_0 in parentheses were derived by another method discussed in a later section.)

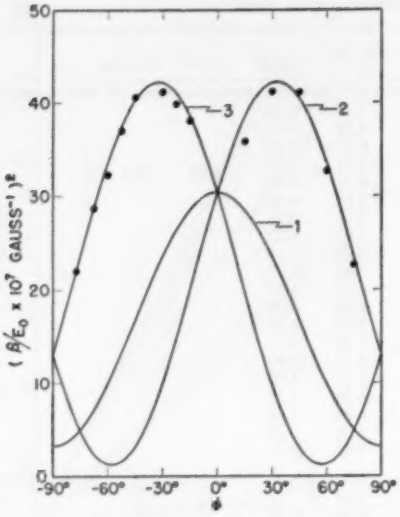


FIG. 11. As 1 (impure). Square of the period for the short-period terms *versus* ϕ for mode (*b*), binary and trigonal axes horizontal.

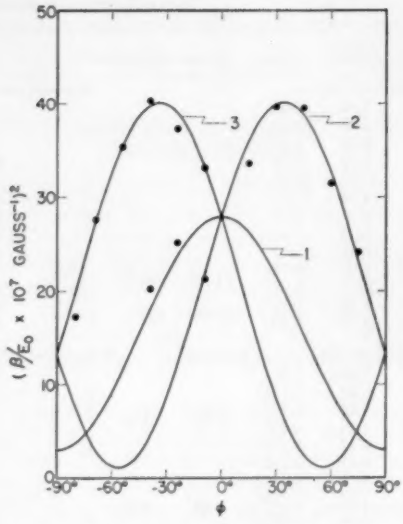


FIG. 13. As 2 (pure). Square of the period for the short-period terms *versus* ϕ for mode (*b*), binary and trigonal axes horizontal. Where two points occur for one angle, beats were observed.

4. Field Dependence of Amplitude

According to a modification of Landau's theory by Dingle,[19] broadening of the electronic energy levels resulting from the presence of impurities has the effect of reducing the amplitude of each term in the summation over p in Eq. (1) by a factor

$$\exp(-2\pi^2 pkx/\beta H),$$

where x has the dimension of temperature and is related to the collision time τ by the expression

$$x=h/2\pi^2 k\tau. \tag{17}$$

It can be shown that if Eq. (1) is valid, $T+x$ is given by the product of $-\beta/2\pi^2 k$ and the slope of the straight line obtained when $\log_e\{aH^{\frac{1}{2}}[1-\exp(-4\pi^2 kT/\beta H)]\}$ is plotted against H^{-1} for fixed T. Several such plots

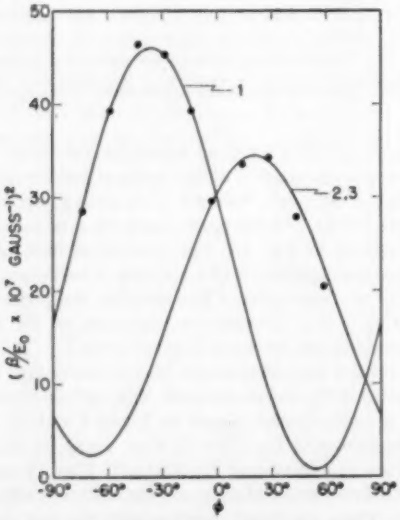


FIG. 12. As 1 (impure). Square of the period for the short-period terms *versus* ϕ for mode (*a*), binary axis vertical, trigonal axis horizontal.

[19] R. B. Dingle, Proc. Roy. Soc. (London) **A211**, 517 (1952).

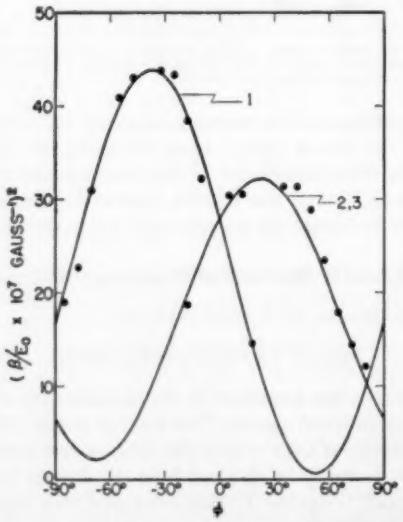


FIG. 14. As 2 (pure). Square of the period for the short-period terms *versus* ϕ for mode (*a*), binary axis vertical, trigonal axis horizontal.

appear in Fig. 16 for the long-period term in As 2. The linearity of these plots verifies the theoretical field dependence of amplitude. Where the plots are nonlinear, it appears that the beats already mentioned prohibit a comparison between theory and experiment. For the indicated fields and temperatures the harmonics were again negligible, although at the lowest temperatures and highest fields their presence was easily recognizable in the long-period term in As 2.

Values of x, derived from plots such as appear in Fig. 16, are tabulated in Tables I and II, and the large magnitudes are in keeping with the known relatively poor crystal quality. The short-period term values are not very reliable because of the relatively small range of fields investigated. In general, x was greater in As 1 as might be expected because of its greater impurity content. Although the dependence of x_l upon ϕ is somewhat irregular, the data for As 2 suggest that x_l increases as ϕ is increased from zero, reaches a maximum value at about $\pi/6$, and then decreases to roughly its initial value as ϕ approaches $\pi/2$. Croft, Love, and Nix[20] have proposed a mechanism which would give rise to an orientation-dependent x, but the scatter of the present data does not justify a detailed interpretation in terms of this mechanism.

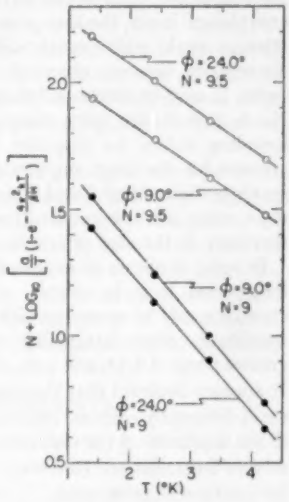


FIG. 15. As 2 (pure). Temperature dependence of the amplitude of the de Haas-van Alphen oscillations for mode (b), binary and trigonal axes horizontal. The open circles correspond to the long-period oscillations, and the solid circles correspond to the short-period terms. Values of β are derived from the slopes of such plots.

5. Absolute Amplitude, Steady Anisotropy, and Number of Electrons

If the degeneracy energy, the electronic effective masses, and x are known for a given de Haas-van Alphen term, the theoretical amplitude can be calculated using Eq. (1) together with Dingle's impurity damping factor. Such a calculation was carried out for the long period term for mode (b) (binary and trigonal axes horizontal) for As 1 with $\phi=25.2°$, $H^{-1}=9\times10^{-6}$, and $T=1.389°$K, and for As 2 with $\phi=24.0°$, $H^{-1}=9\times10^{-6}$, and $T=4.216°$K. The observed amplitudes exceeded the calculated amplitudes in As 1 and As 2 by factors of 28 and 29 respectively. Similar calculations for the short-period term, using long-period term values for x (which is probably not too bad an approximation),

TABLE III. The de Haas-van Alphen parameters for arsenic.[a]

	As 1 (impure)	As 2 (pure)
E_{01} (erg)	1.43×10^{-14}	1.59×10^{-14}
$m_\perp/m_0$	3.18×10^{-3}	3.10×10^{-3}
m_{11}/m_0	0.468 (4.4×10^{3})	0.231 (1.4×10^{3})
n_l per atom	1.8×10^{-4} (6.4×10^{-5})	1.5×10^{-4} (3.6×10^{-5})
δ (radians)	0.50π	0.31π
E_{02} (erg)	29.3×10^{-14}	29.4×10^{-14}
m_1/m_0	0.187	0.193
m_2/m_0	0.956	1.07
m_3/m_0	1.77	1.78
m_4/m_0	−1.25	−1.33
n_s per atom	1.2×10^{-4}	1.2×10^{-4}

[a] Except where double entries appear, the probable errors are of the order of a few percent. The values for m_{11}/m_0 appearing in parentheses were derived from the absolute amplitude of the long-period oscillations. The large discrepancy between these values and the ones obtained as described in part 3 of "Analysis and Evaluation of Parameters" is discussed in the text. n_s refers to each of the three short-period ellipsoids.

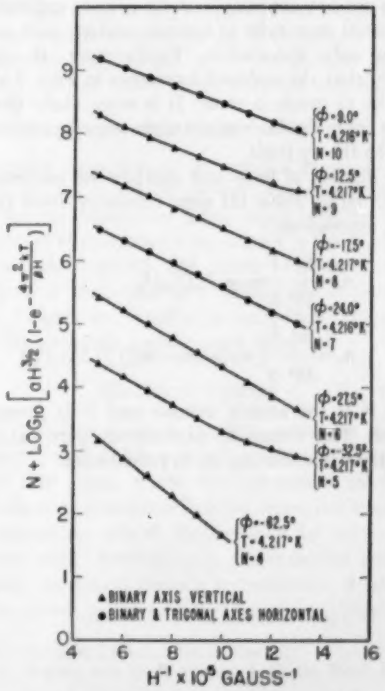


FIG. 16. As 2 (pure). Field dependence of the amplitude of the long-period oscillations. Values of the amplitude damping parameter x are derived from the slopes of such plots if linear.

[20] Croft, Love, and Nix, Phys. Rev. 95, 1403 (1954).

TABLE IV. Steady magnetic anisotropy of arsenic.

T (°K)	$(\chi_{\parallel}-\chi_{\perp})_{\text{steady}}\times10^{6}$ (emu) As 1 (impure)	As 2 (pure)
300	−0.82	−0.54
77	−1.2	−0.80
4.2	−1.3	−0.91

yielded agreement between theory and experiment to within a factor of two in several cases. This is quite satisfactory agreement in view of the uncertainties in the use of the long-period term x. On the other hand, the discrepancy in the long-period case is not easily accounted for. Examination of the parameters involved in the calculations reveals $m_{\parallel}$ (derived from the $\cos^2\phi=0$ intercept in Figs. 9 and 10) as the least accurately determined. Adjustment of $m_{\parallel}$ by an amount sufficient to bring experimental and calculated amplitudes into accord requires the values for $m_{\parallel}/m_0$ appearing in parentheses in Table III. Further possible support for these larger masses exists in that they bring the calculated steady anisotropy [first term in Eq. (1)] into accord (within 10 percent) with the experimentally observed values listed in Table IV. However, even though the effect of the short-period term electrons was also taken into account in this comparison, other conduction electrons (whose de Haas-van Alphen oscillations are not observable under present experimental conditions) may exist in arsenic, making such a comparison only speculative. Furthermore, it appears unlikely that the $\cos^2\phi=0$ intercepts in Figs. 9 and 10 could be so much in error. It is more likely that the answer lies in the inadequacy of the one-ellipsoid scheme or in the theory itself.

The number of long- and short-period electrons per atom listed in Table III were calculated from the following expressions.

$$n_l=\frac{8\pi}{3h^3}\frac{V}{N}m_{\perp}m_{\parallel}{}^{\frac{1}{2}}(2E_{0l})^{\frac{1}{2}}, \tag{18}$$

$$n_s=\frac{8\pi}{3h^3}\frac{V}{N}[m_1(m_2m_3-m_4{}^2)]^{\frac{1}{2}}(2E_{0s})^{\frac{1}{2}}, \tag{19}$$

where V is the atomic volume and N is Avogadro's number. The values in parentheses were calculated using the values for $m_{\parallel}/m_0$ in parentheses.

CONCLUDING REMARKS

The forms of the observed temperature and field dependences of the amplitude of the de Haas-van Alphen oscillations in arsenic are in good accord with theory. However, the phase of the long-period oscillations is at wide variance with theory as is typical in other crystals. In this connection it is of interest to note that As 1, which exhibited a greater maximum period than did As 2, also exhibited the greater phase. Furthermore, in zinc[21] (which displays a temperature-dependent period), as the maximum period increases the phase increases. A generalization on this observation may, however, be premature.

The short-period oscillations in arsenic can be accounted for quite satisfactorily by a three-ellipsoid constant-energy surface scheme. On the other hand, the one-ellipsoid scheme proposed as an approximation for the long-period term is less satisfactory because of the unexplained beats, the long-period oscillations observed (though weak) with the trigonal axis vertical, and the discrepancy between observed and calculated amplitudes. It may be mentioned that the values for $m_{\perp}$ and E_{0l} in arsenic are quite comparable with the corresponding values for graphite, and that in order to account for the large amplitude of the oscillations in graphite, Shoenberg[2] found it necessary to attribute to $m_{\parallel}$ a value of a few hundred free-electron masses as is necessary in the case of arsenic.

In order to obtain an explanation for the beats in the long-period term in arsenic, it is likely that better crystals would be necessary with correspondingly larger amplitude, thus facilitating measurements over a greater range of fields and with the trigonal axis vertical. It appears doubtful that the observed beats arise from small differently-oriented regions in the crystals in view of the similarity of the results for the two crystals used in this investigation. However, this possibility cannot be ruled out with certainty.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the generous assistance of Mr. Ralph Anonsen in the acquisition and reduction of data.

[21] Ted G. Berlincourt and M. C. Steele, Phys. Rev. **95**, 1421 (1954).

Infrared Spectra of Ferrites*

R. D. WALDRON
Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received April 11, 1955)

The infrared spectra of 7 ferrites of the formula MFe_2O_4, where M designates a divalent metal, are presented and analyzed. Electronic absorption was observed in the visible and near-infrared regions. Two absorption bands arising from interatomic vibrations were measured and force constants calculated for the stretching of bonds between octahedral or tetrahedral metal ions and oxide ions. These force constants are in agreement with the elastic and thermodynamic properties of these compounds and are sensitive to distribution of metal ions between the alternate sites. The integrated vibrational band intensities were measured: they are compatible with predominantly ionic bonding for these structures.

STUDIES of the relations between structure and the electromagnetic response of ferromagnetic semiconductors are useful in understanding their properties. Since the electric and magnetic properties of these materials are decisively dependent on the precise configuration of the atoms or ions in these structures, methods of nondestructive analysis are especially suited to such investigations. In particular, the vibrational, electronic, and magnetic dipole spectra can give information about the position and valence of the ions in the crystal lattices.

In the present work, the infrared spectra of seven ferrites with the general formula MFe_2O_4, where M represents a divalent metal ion, were examined in the frequency range 280–4000 cm^{-1}; in several cases the investigation was carried up to the visible range. No previous infrared studies of these materials were found in the literature with the exception of the reflection spectrum of magnetite in the near infrared measured by Coblentz.[1]

EXPERIMENTAL

The spectra were recorded on a Beckman IR-3 spectrophotometer equipped with CaF_2, KBr, and KRS-5 prisms. Calibration was made by recording H_2O, NH_3, CO_2, CH_4, HCl, and HBr lines previously measured with considerable accuracy on grating instruments.[2]

The samples were made according to the technique of Stimson[3] and Schiedt[4]; 2.5 mg of the ferrite was mixed with 0.8 g of powdered KBr, and placed in a cylindrical die of 20 mm diam. The die was evacuated (0.1 mm Hg) for about 30 min to insure dryness and the sample then pressed for about 10 min at 30 tons/inch^2. Clear disks of approximately 1 mm thickness were obtained with usable transmission to 280 cm^{-1}. The ferrites themselves were prepared by precipitation and thermal decomposition of the corresponding oxalates as described by Wickham.[5]

Measurements were made at room temperature and at liquid nitrogen temperatures in a Dewar cell with AgCl windows. The cell construction was essentially the same as described by Wagner and Hornig.[6]

DATA

Spectra of MFe_2O_4 in the range 280–2500 cm^{-1} are shown in Fig. 1; M designates Co, Fe, Mg, Mn, Ni, or Zn. The spectrum of the mixed ferrite $Ni_{0.3}Zn_{0.7}Fe_2O_4$, the low-temperature spectrum of Fe_2O_4, and the reflection spectrum of $NiFe_2O_4$ are presented in Fig. 2. Low-temperature spectra of the other simple ferrites were examined above 430 cm^{-1}, but differed only in minor detail from the room temperature spectra.

ANALYSIS

A cursory inspection of the spectra shows two absorption bands below 1000 cm^{-1} as a common feature of all the ferrites. Absorption in this region is not restricted to this class of compounds but occurs in the spectra of most metallic oxides.[7] The bands arise from lattice vibrations of the oxide ions against the cations. At higher frequencies, gradually increasing absorption caused by electronic transitions is observed.

Electronic Spectra

The high-frequency absorption possesses a low-frequency cutoff which cannot be defined without ambiguity. For cases where the absorption reaches a maximum or approaches a limiting value, the frequency of ½ maximum optical density may be taken as a reference point. Alternatively, the inflection point in the graph of optical density *vs* frequency, if present, may be chosen as the reference frequency. This latter

* Sponsored by the Office of Naval Research, the Army Signal Corps, and the Air Force.

[1] W. W. Coblentz, *Additional Investigations of Infrared Spectra* (Carnegie Institution, Washington, D. C., 1908).

[2] Downie, Magoon, Purcell, and Crawford, J. Opt. Soc. Am. **43**, 941 (1953).

[3] M. M. Stimson and M. J. O'Donnell, J. Am. Chem. Soc. **74**, 1805 (1952).

[4] U. Schiedt and H. Reinwein, Z. Naturforsch. **7b**, 270 (1952); Appl. Spect. **7**, 75 (1953).

[5] D. G. Wickham (personal communication).

[6] E. L. Wagner and D. F. Hornig, J. Chem. Phys. **18**, 296 (1950).

[7] Bands in this region have been observed in Al_2O_3, MgO, Fe_2O_3, TiO_2, $BaTiO_3$, etc. See, for example, L. Harris, J. Opt. Soc. Am. **45**, 27 (1955); F. Matossi, J. Chem. Phys. **19**, 1543 (1951).

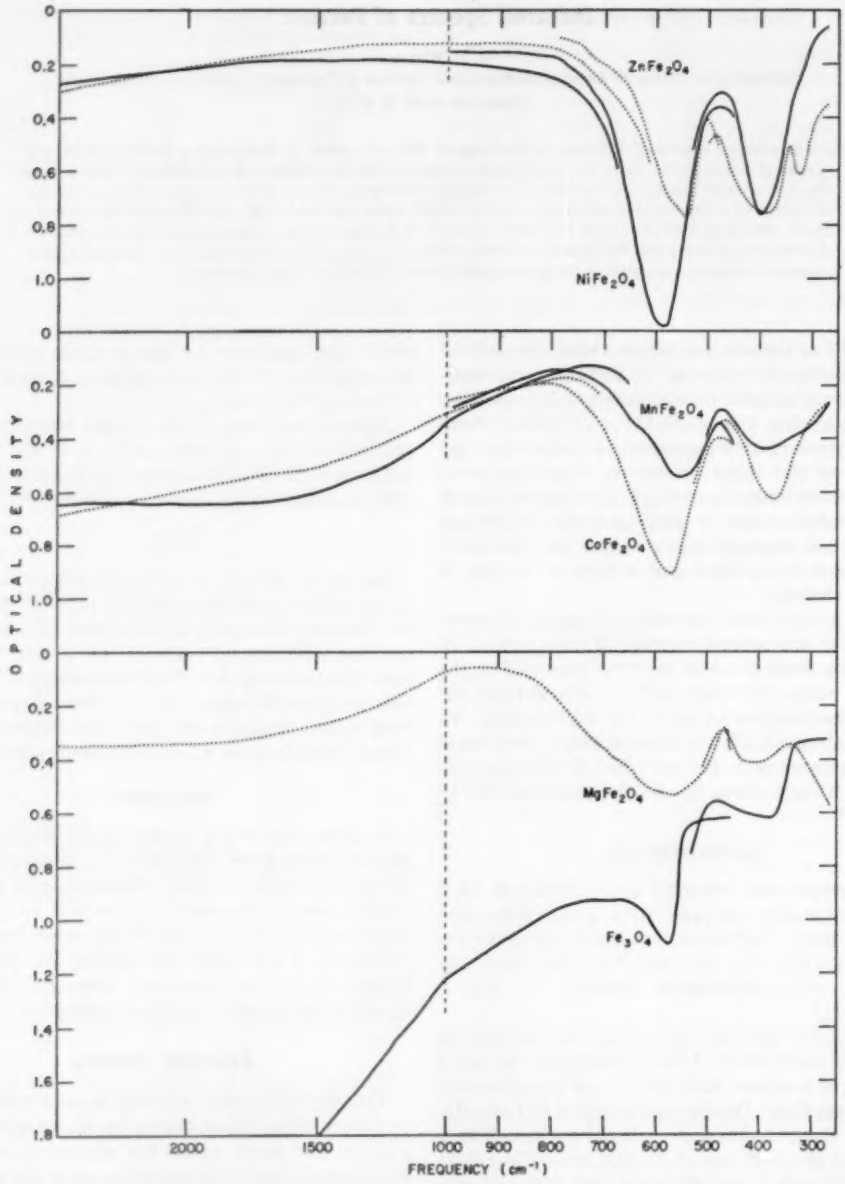


FIG. 1. Infrared absorption spectra of simple ferrites. Optical density measured relative to standard KBr disk. Sample weight: 2.50 mg (Fe 5.00 mg); sample area: 3.14 cm². Multiple lines indicate alternate runs or prism changes.

criterion will be used except where noted. Table I shows these threshold values for the electronic transition. These data are subject to considerable uncertainty because that portion of the scattering caused by the ferrite particles cannot be distinguished from absorption losses.

If these absorptions represent transitions to a conduction band of the crystal, the tabulated values can be expressed as activation energies. Some of the electronic bands give an indication of structure at higher frequencies; the scattering losses of our samples preclude their positive identification.

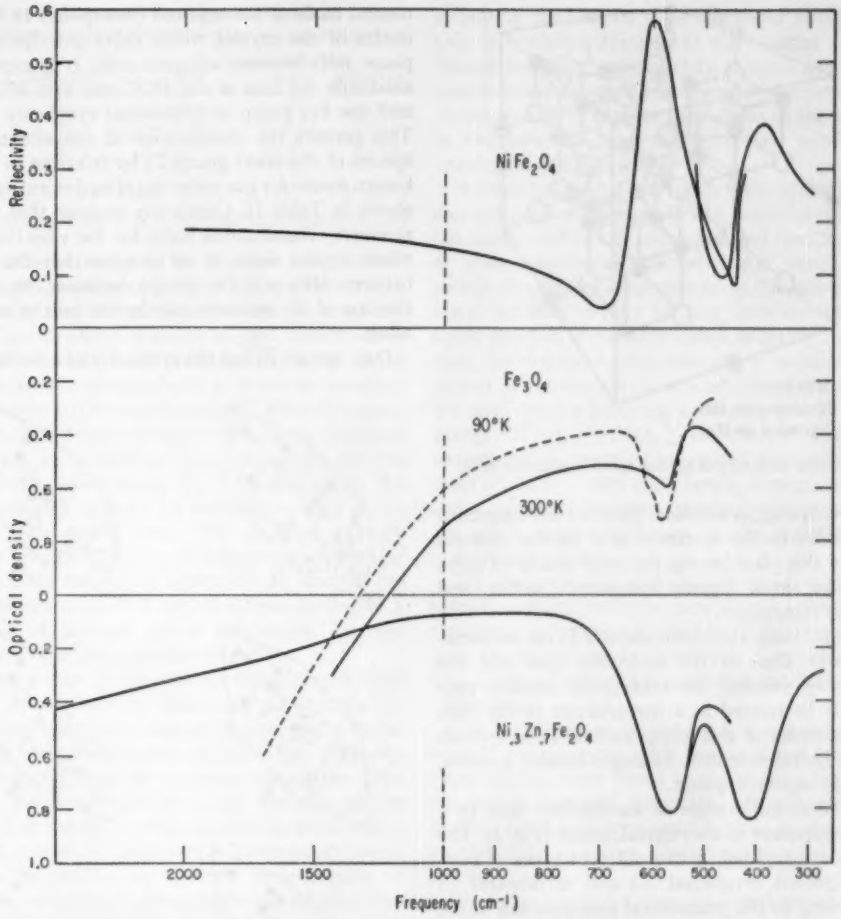


FIG. 2. Special infrared spectra of ferrites. Top: reflection spectrum of single crystal of $NiFe_2O_4$ (normal incidence [110] face). Center: low-temperature absorption spectrum of Fe_3O_4 (2.50 mg). See legend of Fig. 1. Bottom: absorption spectrum of mixed ferrite $Ni_{0.3}Zn_{0.7}Fe_2O_4$. See legend of Fig. 1.

Vibrational Spectra

The bands in the 300 to 700 cm^{-1} region are assigned to the fundamental vibrations of the ions of the crystal lattice. For the analysis of such spectra, it is necessary to consider the vibrational spectrum of periodic structures. The vibrational problem is most conveniently treated by classification of crystals according to the continuity of bonding as (1) continuously bonded; (2) discontinuously bonded; and (3) intermediate. In continuously bonded crystals, the atoms are bonded to all nearest neighbors by equivalent forces (ionic, covalent, or van der Waals) and the frequency distribution of vibrations given by a Debye or Born-von Karman treatment of the classical mechanical problem. In this class, we include simple ionic crystals such as the alkali halides; diamond and its homologues; metals; and rare-gas crystals. In the discontuously bonded or molecular crystal, sets of atoms are tightly bound by (intramolecular) chemical valence forces and separated from adjacent sets by weak (intermolecular) van der Waals forces. In a first approximation, the spectrum resembles that of a gas molecule with the vibrational lines slightly broadened due to weak inter-

TABLE I. Threshold frequency of the electronic absorption band.

Compound	cm^{-1}	ev
$CoFe_2O_4$	900±50	0.11
$FeFe_2O_4$	1200±200	0.15
$MgFe_2O_4$	1060±50	0.13
$MnFe_2O_4$	930±50	0.12
$NiFe_2O_4$*	2700±600	0.33
$ZnFe_2O_4$	2300±200	0.29

* Inflection point ill-defined due to nearly constant slope.

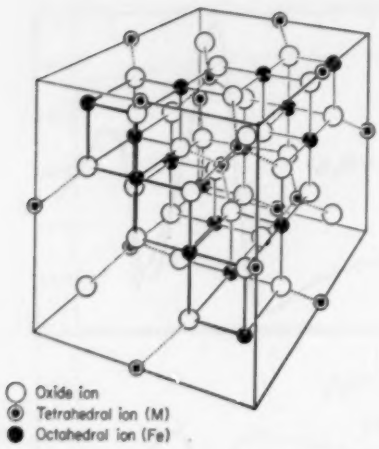


FIG. 3. Crystal structure of normal spinel (cubic unit cell).

molecular coupling; in addition, there is a low-frequency branch similar to the spectrum of a heavier rare-gas crystal. To this class belong the solid phases of polyatomic gases, most organic compounds, sulfur, and nonmetallic compounds.

In the third class, the intermolecular forces are somewhat greater than in the molecular case and the branches may overlap; the vibrational problem may occasionally be treated as a perturbation of the class 2 case. Examples of this group include ionic crystals containing polyatomic ions, hydrogen-bonded crystals, and strongly dipolar crystals.

X-ray diffraction studies of the ferrites[8] lead to a structure equivalent to the mineral, spinel (Fig. 3). The metal ions are situated in two different types of positions, designated octahedral (o) and tetrahedral (t) sites according to the geometrical configuration of the oxygen nearest neighbors. Some doubt exists, however, concerning the distribution of the different metal ions between the alternate sites. No isolated molecular groupings occur in the structure, so we assign the ferrites to the class 1 or the continuous case.

The complete frequency distribution consists of $42N$ vibrational modes, where N is the number of primitive unit cells (PUC) in the crystallite (3 degrees of freedom for each ion; 14 ions per PUC). The selection rules require that only modes of proper symmetry absorb infrared radiation. The space group symmetry for a normal spin is $O_h^7(F\,4_1/d\,\bar{3}\,2/m)$, with 14 atoms per primitive unit cell ($M_2Fe_4O_4$). (The term normal or inverse spinel designates a structure in which M ions are situated on the tetrahedral or octahedral sites, respectively.)

The complete vibrational spectrum can be classified in terms of the vibrations of a single unit cell. Each normal mode of the unit cell corresponds to N normal modes of the crystal which differ principally in the phase shift between adjacent cells. It is expedient to subdivide the ions of the PUC into two MO_4 groups and one Fe_4 group of tetrahedral symmetry (Fig. 4). This permits the classification of the vibrations into species of the point group T_d by referring to the well-known results for gas molecules of equivalent symmetry[9] shown in Table II. Continuity requires that the same symmetry classification holds for the vibrations of the whole crystal since, if we imagine that the coupling between MO_4 and Fe_4 groups decreases, we approach the case of $3N$ gas molecules in the limit of zero interation.

Only species F_2 has the symmetry of a vector set and

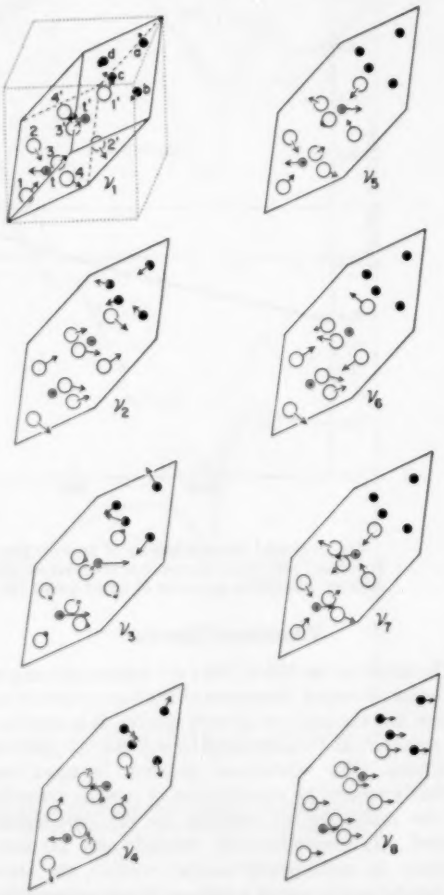


FIG. 4. Species F_2 normal vibrations of a spinel (schematic). Cubic unit cell (dotted line) (see Fig. 3). Rhombohedral primitive unit cell PUC (solid line). Tetrahedral groups $MO_4(t,1,2,3,4)$, $MO_4'(t',1',2',3',4')$, $Fe_4(a,b,c,d)$.

[8] R. W. G. Wyckoff, *Crystal Structures* (Interscience Publishers, Inc., New York, 1951), Vol. II, Chap. 8.

[9] G. Herzberg, *Infrared and Raman Spectra* (D. Van Nostrand Company, Inc., New York, 1945), pp. 99 ff.

TABLE II. Symmetry classification of vibrations.

T_d species	Degeneracy	Degrees of freedom MO_4	Fe_4	$2MO_4+Fe_4$
A_1	(1)	(1)×1	(1)×1	(1)×3
E	(2)	(2)×1	(2)×1	(2)×3
F_1	(3)	(3)×1	(3)×1	(3)×3
F_2	(3)	(3)×3	(3)×2	(3)×8
	Total	15	12	42

hence is infrared active. The selection rules further require that the phase shift between successive cells match that of the electromagnetic wave of equal frequency. Since the ratio of sound to light velocity is so low, this phase shift is extremely small and the active mode effectively corresponds to a vibration invariant with respect to lattice translations. Of the $42N$ normal vibrations, $24N$ belong to species F_2, 24 are invariant with respect to translation, and only eight discrete infrared bands should occur. In an inverse spinel, the local symmetry is lowered by substitution of $\frac{1}{4}$ of the octahedral sites by M ions. This splits each triply degenerate vibration into three vibrations of different frequency. Disorder in the lattice would destroy the translational invariance of the vibrations leading to a distribution of infrared active frequencies. Isotopic substitution would also produce this effect.

The precise form of the normal vibrations depends upon the potential energy field except where the requirements of symmetry uniquely determine the mode. Of the eight possible frequencies, three are internally inactive and one is the zero frequency translation. (The internally inactive vibrations arise because of the equivalence of the MO_4 groups and the location of the Fe ions on centers of symmetry.) The normal coordinates of species F_2 are pictured schematically in Fig. 4. Each oxide ion is bonded to three octahedral ions and one tetrahedral ion. In the normal spinel, the three octahedral bonds lie along the coordinate axes and provide an isotropic force field which would permit the oxide ion to oscillate in any direction with a constant frequency if the tetrahedral bonds were absent. The tetrahedral bonds have the effect of substantially increasing the frequency for vibrations along the $t-O$ axis (ν_1), while vibrations at right angles (ν_2) would change frequency only slightly if small bending force constants are assumed. The two low-frequency active modes, ν_3 and ν_4, involve oscillations of the metal ions in the isotropic force fields of their octahedral or tetrahedral environments.

Since only two inrared bands were accessible in the spectral range covered, only two potential constants of the energy field can be calculated, and the identification of these constants with particular bonds requires certain simplifying assumptions. In the following treatment, we shall neglect bending force constants: this approximation is probably closer to reality in the ionic then the molecular case, where it is occasionally employed in the evaluation of stretching force constants. We shall use the following expression for the potential energy: $2V=k_t\sum r_t^2+k_o\sum r_o^2$, where k_t and k_o designate the force constants associated with unit displacement of the $t-O$ and $o-O$ bonds, respectively, and r_t and r_o represent the corresponding components of displacement from equilibrium in the direction of the appropriate bonds.

The determination of the vibrational frequencies and normal coordinates can be carried out by the method of symmetry coordinates devised by Wilson[10]; but in our case, the modes can be determined directly by inspection. Table III gives the coordinates for the four active vibrations.

The zero frequency result for ω_4 is not unexpected in view of the potential field chosen: if the MO_4 groups were isolated, there would be no restoring force for the degenerate bending vibration. Introduction of a bending-force constant would lead to simultaneous equations for ω_2 and ω_4 similar to the equations for ω_1 and ω_3.

We assign the high-frequency line observed to ν_1 and the low-frequency line to ν_2 for the following reasons: ν_2 has a lower frequency than ν_1 as was discussed qualitatively; the frequencies are about those expected for roughly equivalent force constants (if $k_t=k_o$, $\nu_1/\nu_2\simeq\sqrt{2}$); and the intensities are nearly equal, as expected if the motions are predominantly those of the oxide ion. The

TABLE III. Normal coordinates of infrared active vibrations.

ν_1, ν_2		ν_3		ν_4
$a=u(-1\ \ 1\ \ 1)$	$1=1\ \ 1\ \ 1$	$a=U(-2\ -1\ -1)$	$1=2\ -1\ -1$	$a,\ 2=\ x\ \ 1\ \ 1$
$b=u(-1\ -1\ -1)$	$2=1\ -1\ -1$	$b=U(-2\ \ 1\ \ 1)$	$2=2\ \ 1\ \ 1$	$b,\ 1=\ x\ -1\ -1$
$c=u(-1\ -1\ \ 1)$	$3=1\ -1\ \ 1$	$c=U(-2\ \ 1\ -1)$	$3=2\ \ 1\ -1$	$c,\ 4=\ x\ -1\ \ 1$
$d=u(-1\ \ 1\ -1)$	$4=1\ \ 1\ -1$	$d=U(-2\ -1\ \ 1)$	$4=2\ -1\ \ 1$	$d,\ 3=\ x\ \ 1\ -1$
$t=v(-1\ 0\ 0)$		$t\equiv 0$		$t=-X\ \ 0\ \ 0$
$m_t v+2m_o u=4m_0=64$		$m_o U=2m_0=32$		$m_t X=(2m_o+4m_0)x$
$\omega_1,_2^2=4k_t\frac{(1/v+\frac{1}{4})}{m_t}$		$\omega_3^2=2k_o\frac{(1/U+1)}{m_o}$		$X=2-x$
$=2k_o\frac{(1/u+1)}{m_o}$				$\omega_4\equiv 0$

where $1=1'$, $2=2'$, etc., $t=t'$, $\omega=2\pi\nu$.

[10] J. B. Howard and E. B. Wilson, Jr., J. Chem. Phys. 2, 630 (1934).

TABLE IV. Infrared spectral frequencies of ferrites.

Compound	ν_1	ν_2		Other ν's (cm^{-1})
$CoFe_2O_4$	573	374		shoulder *ca* 320
$FeFe_2O_4$	570 {579[a] 610[a]}	370–80		
$MgFe_2O_4$	565	406	{380 430–40}	shoulder *ca* 735
$MnFe_2O_4$	550	392		shoulder *ca* 645, 320
$NiFe_2O_4$	587	396		
$ZnFe_2O_4$	555 {540 575}	393		shoulder *ca* 325
$Ni_{0.3}Zn_{0.7}Fe_2O_4$	563	409		

[a] Liquid N_2.

last fact rules out overtones or combinations of lower frequency vibrations as the source of one of the absorption bands. The presence of weaker bands and line broadening in this region may be due to splitting of degeneracies, overtone or combination bands and precise assignment is not possible.

Table IV gives the measured frequencies of the vibrational bands of the ferrites. The customary narrowing of bands was not observed at liquid N_2 temperature, which may indicate splitting of degeneracies or disorder, but no significant trend from normal to inverse spinels was apparent.

The force constants were computed as follows: ν_2 was used to determine k_a: three values were obtained corresponding to $m_a = Fe$, M, and $(Fe+M)/2$, respectively. Next, ν_1 and each of the values of k_a were used to determine u and v; the corresponding values of m_t were M, Fe, and Fe, respectively. Finally, substitution in the alternate frequency equation gave k_t. Table V summarizes the results.

The first column of each section corresponds to a normal spinel, while the next two columns represent inverse structures. The mixed ferrite was calculated with the masses indicated. For $MgFe_2O_4$, the disparity in masses becomes serious, and calculations based on subdividing the low-frequency band into several components is necessary. The accompanying spectrum is considerably broadened with evidence of multiplet structure. The assigned values are considered to represent the actual constants for reasons discussed below.

DISCUSSION

Of the three ferrites believed to be predominantly inverse, i.e., Co, Fe, and Ni,[11] with tetrahedral sites occupied by Fe, the constants k_t are very nearly equal (*ca* 1.66×10^5 dynes/cm). For $ZnFe_2O_4$, which is a normal spinel, as well as Mg and Mn in which some doubt exists as to the configuration,[11] the values of k_t are substantially lower. This phenomenon, if generally true, might assist in the elucidation of the structure of complex ferrites.

One would predict a lowering of the force constant for either site if the mean ionic charge for such a site were lowered. The lowered electrostatic energy would lead to a slight increase of the $t-O$ bond length and a reduction in the repulsive forces between the ions leading to a lower force constant. In the case of Mg, it is likely that some divalent Mg ions are situated on tetrahedral sites which would be consistent with the hypothesis advanced above, and would explain the broad high-frequency shoulder occurring in the infrared spectrum. For $MnFe_2O_4$, one would expect that either Mn^{++} ions occupy some tetrahedral sites, or that Fe^{++} occurs in some tetrahedral sites, and an equivalent amount of Mn^{+++} or Mn^{++++} is present in the lattice. The former case would represent a tendency to a normal spinel structure, while the latter would be atomically inverse but electrically normal. Either situation would not affect the saturation magnetic moment of the crystal.[11]

The valence bond-force constants may be used to calculate the elastic constants of the crystal, and they in turn are useful in the evaluation of the compressibility and heat capacity. For this calculation, it is convenient to compute the energy required to compress

TABLE V. Force constants of the ferrites.

Compound	k_a				k_t			
	$m_a = Fe$	M	$\frac{1}{2}(Fe+M)$	Assigned	$m_t = M$	Fe	Fe	Assigned
$CoFe_2O_4$	0.838	0.854	0.846	0.85	1.705	1.677	1.645	1.66
$FeFe_2O_4$	{0.820[a] 0.865[a]}			0.84		{1.66[a] 1.632[a]}		1.65
$MgFe_2O_4$	0.988[b] 0.865[a]	{0.752[d] 0.788[d]}	0.864[b]	0.86	1.178[b] 1.253[a]	{1.521[d] 1.469[d]}	1.522[b]	1.52
$MnFe_2O_4$	0.920	0.915	0.917	0.92	1.394	1.402	1.402	1.40
$NiFe_2O_4$	0.939	0.957	0.948	0.95	1.715	1.69	1.66	1.67
$ZnFe_2O_4$	0.925			0.925	{1.422[e] 1.542[e]}			1.477
$Ni_{0.3}Zn_{0.7}Fe_2O_4$	$m_a = (0.85\ Fe+0.15\ Ni)$ $m_t = (0.3\ Fe+0.7\ Zn)$			1.005				1.455

[a] Calculations based on: $\nu_2 = 370$–380 cm^{-1}.
[b] Calculations based on: $\nu_2 = 406$ cm^{-1}.
[c] Calculations based on: $\nu_2 = 380$ cm^{-1}.
[d] Calculations based on: $\nu_2 = 430$–440 cm^{-1}.
[e] Calculations based on: $\nu_1 = 540$–575 cm^{-1}.

[11] E. J. W. Verwey and E. L. Heilmann, J. Chem. Phys. **15**, 174 (1947).

TABLE VI. Elastic constants of the ferrites.

Ferrite	r_0	C_1	$x_0=r_0/3$	$C_1=C_2/9$	$c_{11}=c_{12}=1/\beta$
$CoFe_2O_4$	0.3305	7.932	0.1102	0.8813	1.794×10^{12} dynes/cm²
$FeFe_2O_4$	0.3305	7.932	0.1102	0.8813	1.772×10^{12} dynes/cm²
$NiFe_2O_4$	0.3188	7.656	0.1063	0.8506	1.944×10^{12} dynes/cm²

the unit cell a standard distance: A, in three dimensions; and B, along one axis. The unit-cell edge a_0 is determined by two $o-O$ bonds parallel to a_0 and two $t-O$ bonds at $\cos^{-1}(1/\sqrt{3})$ with a_0.

For case A, unit displacement requires that $2r_t/\sqrt{3}+2r_o=1$. Since the unit cell contains 32 $t-O$ bonds and 96 $o-O$ bonds, we may write:

$$V_3=(32/2)k_t r_t^2+(96/2)k_o r_o^2$$

as the potential energy per unit cell for 3-dimensional compression. If we set $k_t=\gamma k_o$ and substitute to eliminate r_t, we obtain:

$$V_3=12k_o[\gamma-4\gamma r_o+4(1+\gamma)r_o^2]=C_3(\gamma)k_o$$

for unit displacement.

To determine the individual displacements yielding the lowest energy, we set:

$$\partial V_3/\partial r_o=0,\quad A[-4\gamma+8(1+\gamma)r_o]=0,\quad r_o=\gamma/2(1+\gamma).$$

For an arbitrary displacement,

$$V_3=C_3k_o(a-a_0)^2=C_3k_oa_0^2\alpha^2,$$

where $\alpha=(a-a_0)/a_0$. Dividing by the unit-cell volume a_0^3, we finally obtain the expression for the potential energy per unit volume:

$$\bar{V}_3=C_3k_o\alpha^2/a_0.$$

Elasticity theory gives the following expression for the energy density:

$$\bar{V}=\tfrac{1}{2}\sum_{i,j=1}^{6}c_{ij}\alpha_i\alpha_j.^{12}$$

In the special case of uniform compression of a cubic crystal, we have $\bar{V}_3=\frac{3}{2}(c_{11}+2c_{12})\alpha^2$. It follows that $c_{11}+2c_{12}=2C_3k_o/3a_0$.

For case B, we resolve r_t into components x_t, y_t, and z_t, and subdivide the displacements of the 96 $o-O$ bonds into 32 x_o, 32 y_o, and 32 z_o. By symmetry, $y_t=z_t$, $r_t=(x_t+2y_t)/\sqrt{3}$, and for unit displacement along the x axis and zero displacements along the y and z axes, we have:

$$\begin{aligned}&2x_t+2x_o=1,\\&2y_t+2y_o=0,\end{aligned}\qquad V_1=16k_t\left(\frac{x_t+2y_t}{\sqrt{3}}\right)^2+16k_o(x_o^2+2y_o^2).$$

Substituting to eliminate x_t, y_t, and k_t yields:

$$V_1=(4k_o/3)(\gamma-4\gamma x_o-8\gamma y_o+4\gamma x_o^2+16\gamma x_oy_o+16\gamma y_o^2+12x_o^2+24y_o^2)=C_1(\gamma)k_o$$

for unit displacement. To minimize the energy, we set:

$$\partial V_1/\partial x_o=0,\quad A(-4\gamma+8\gamma x_o+16\gamma y_o+24x_o)=0,$$
$$\partial V_1/\partial y_o=0,\quad A(-8\gamma+16\gamma x_o+32\gamma y_o+48y_o)=0.$$

Solving for x_o and y_o gives:

$$x_o=\frac{\begin{vmatrix}\gamma & 4\gamma\\ \gamma & 6+4\gamma\end{vmatrix}}{\begin{vmatrix}6+2\gamma & 4\gamma\\ 2\gamma & 6+4\gamma\end{vmatrix}},\qquad y_o=\frac{\begin{vmatrix}6+2\gamma & \gamma\\ 2\gamma & \gamma\end{vmatrix}}{\begin{vmatrix}6+2\gamma & 4\gamma\\ 2\gamma & 6+4\gamma\end{vmatrix}},$$

or $x_o=y_o=\gamma/6(1+\gamma)$. For arbitrary displacements, $V_1=C_1k_oa_0^2\alpha^2$ and $\bar{V}_1=C_1k_o\alpha^2/a_0$.

From elasticity theory: $\bar{V}_1=\frac{1}{2}c_{11}\alpha^2$. Thus $c_{11}=2C_1k_o/a_0$, and using the results obtained for case A, we obtain $c_{12}=(C_3/3-C_1)k_o/a_0$.

The compressibility is defined by the expression: $\beta=-\partial\bar{V}/\bar{V}\partial p$, where $\bar{V}$ denotes the volume of the crystal. It can be shown that $1/\beta=(c_{11}+2c_{12})/3=2C_3k_o/9a_0$ for cubic crystals. Table VI gives the computed constants and compressibilities for Co, Fe, and $NiFe_2O_4$.

Experimental values reported for $1/\beta$ for magnetite range from 1.61 to 1.78×10^{12} dynes/cm², with measurements of $c_{11}=2.70$ to 2.75×10^{12} dynes/cm² and $c_{12}=1.04$ to 1.08×10^{12} dynes/cm² also quoted.[13]

The agreement with experiment is quite satisfactory for the compressibility. The discrepancy with the elastic constants can be traced to the neglect of the bending-force constants. In fact, in the approximation used for the potential field, there would be no resistance to one-dimensional compression if the lateral dimensions were free to expand (analogous to the zero frequency motion for ν_4). This leads to a value of c_{11} which is too low, and c_{12} is accordingly too high since it is derived by subtraction of a one-dimensional term from the three-dimensional term. The equality of c_{11} and c_{12} seems to be fortuitous.

The molar heat capacity at constant volume, C_v, may be calculated in the Debye theory approximation either by using the computed elastic constants or by using the limiting frequencies observed in the infrared spectrum.

In the former case, the elastic wave velocities are

[12] F. M. Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), pp. 94 ff.

[13] M. S. Doraiswami, Proc. Indian Acad. Sci. A25, 413 (1947).

TABLE VII. Molar heat capacity of magnetite (Fe_3O_4).

T°K	C_p[a]	C_v[b]	C_v (elas. const)	C_v (IR freq.) cal/mole
60.5	5.495	5.43	2.166	6.094
70.5	7.331	7.25	3.359	8.128
100.7	13.67	13.56	8.369	13.85
127.5	19.16	19.02	13.46	18.26
153.7	22.63	22.46	18.08	22.05
191.4	26.98	26.77	23.54	26.34
235.7	31.18	30.92	28.14	30.05
261.1	32.68	32.39	30.14	31.67
299.7	34.39	34.06	32.47	33.60

[a] R. W. Millar, J. Am. Chem. Soc. 51, 215 (1929).
[b] $C_p - C_v = 0.0011T$ assumed.

given by: $v_l = (c_{11}/\rho)^{\frac{1}{2}}$, $v_t = [(c_{11}+c_{12})/2\rho]^{\frac{1}{2}}$, where v_l and v_t represent longitudinal and transverse wave velocities and ρ is density.

Using the x-ray computed density for $FeFe_2O_4$, 5.238, and the previously derived elastic constants, we arrive at the following values: $v_l = v_t = 5.816\times10^5$ cm/sec. From the relation $\bar{\nu}_{max} = v\sigma_{max}$, we may calculate the high-frequency cutoff of the elastic waves. We may assume that the shortest wave corresponds to $\frac{1}{2}$ wave per mean interatomic distance $\bar{d}$. Since $\bar{d} = \frac{1}{4}(3r_o + r_i) = 7.732\ a_0/32$, we obtain $\sigma_{max} = 2.07/a_0$ and $\bar{\nu}_{max} = 1.438\times10^{13}$ sec^{-1}. This corresponds to a frequency of 480 cm^{-1} which represents an approximate mean between the cut-off frequencies of the oxide and metal ion vibrations, ν_1 and ν_3. The characteristic Debye temperature is defined by $\theta = hc\nu_{max}/k = 1.438\nu_{max} = 690°$. This would give $C_v = 21Rf(T/690)$, where R is the gas constant and f is the Debye function.

In the second method, we may employ a two-parameter treatment by using the frequencies ν_1 and ν_3 as cut-off frequencies for oxide vibrations and metal ion vibrations, respectively. ν_1 was observed directly and ν_3 was calculated by solving the simultaneous equations for u and v in Table III. The results are: $\nu_1 = 570$ cm^{-1}, $u = 10.22$, $v = -19.28$, $\nu_3 = 237$ cm^{-1}, $\theta_1 = 820°$, and $\theta_3 = 341°$. Finally $C_v = R[12f(T/820) + 9f(T/341)]$.

The calculated molar heat capacity is compared in Table VII with the experimental values using a $C_p - C_v$ correction calculated from the experimental values of the thermal expansion coefficient and compressibility at room temperature and assuming the function to be linear in temperature.

The agreement is fairly good, especially with the *IR* frequency method. An improved fit could be obtained by empirically choosing a lower cut-off frequency in the elastic constant method.

INTENSITIES

The measurement of intensities of infrared bands can give useful information concerning the change of electric dipole moment with interatomic distance. For this purpose, the intensities of both ν_1 and ν_2 were measured.

We define the extinction coefficient k by the equation:

$$I_{trans}/I_{incid} = e^{-kx} = 10^{-Ax},$$

where I is the intensity (power) of the radiation, x is the path length, and A is the absorption coefficient. It can be shown that[9]

$$\bar{k} = (8\pi^3 N\nu/3ch)([M]^{n,m})^2$$

for gas molecules, where $\bar{k} = \int k(\nu)d\nu$, N is the number of molecules/cm^3, ν is the frequency, c is the velocity of light, h is Planck's constant, and $[M]^{n,m} = \int \psi_n \mu \psi_m^* d\tau$ is the matrix element corresponding to the electric dipole transition moment between the states m and n. We may expand the transition moment in a power series in the normal coordinate q:

$$[M]^{n,m} = M_0 \int \psi_n \psi_m^* d\tau + \frac{dM}{dq}\int \psi_n q \psi_m^* d\tau + \frac{1}{2}\frac{d^2M}{dq^2}\int \psi_n q^2 \psi_m^* d\tau + \cdots.$$

The first term vanishes by virtue of the orthogonality of the wave functions, while the third term is zero in the harmonic oscillator approximation for fundamental absorption bands. For harmonic oscillators, the integral

$$q^{0,1} = \int \psi_n q \psi_m^* d\tau = 1/(2\alpha)^{\frac{1}{2}},$$

where $\alpha = 4\pi^2 mc\nu/h$, where m is the reduced mass. Thus:

$$M_q^2 = \left(\frac{dM}{dq}\right)^2 = \frac{3\alpha ck}{4\pi^3 N}\bar{k} = \log_e\left(10\frac{3mc^2}{\pi Nx}\right)\cdot \bar{A}x = 1.975\times10^{21}\frac{m\bar{A}x}{Nx}.$$

TABLE VIII. Intensities and derived constants of ferrites.

Compound	Fundamental	$\bar{A}x$	Nx (cm^{-2}) $\times10^{18}$	m (g) $\times10^{-23}$	Mq (esu) $\times10^{-9}$	Q(e)
$CoFe_2O_4$	ν_1	108	2.1	2.14	1.477	1.78
	ν_2	53.5		1.69	0.922	1.11
$FeFe_2O_4$	ν_1	22	2.14	2.16	0.663	0.80
	ν_3	10.65		1.69	0.408	0.49
$MgFe_2O_4$	ν_1	120.8	2.47	2.2	1.46	1.76
	ν_2	47.8		1.69	0.804	0.97
$MnFe_2O_4$	ν_1	66	2.14	2.16	1.15	1.38
	ν_3	56.5		1.69	0.939	1.13
$NiFe_2O_4$	ν_1	140.5	2.1	2.14	1.683	2.02
	ν_3	55.8		1.69	0.942	1.13
$ZnFe_2O_4$	ν_1	88	2.05	2.2	1.37	1.65
	ν_3	91.6		1.69	1.22	1.47

Table VIII gives the measured values for the compounds studied. The last column is defined by the equation $Q=M_q/\sqrt{3}\epsilon$, where ϵ is the electronic charge. Thus Q represents the effective charge in electrons equivalent to the measured intensities. (Since the vibrations are triply degenerate, M_q represents the vector sum of the transition moments for the individual components and the factor $1/\sqrt{3}$ must be incorporated to yield the correct value for a single axis. The effect of refractive index was neglected.)

It may be shown by a semiclassical calculation that the polarization per unit volume and hence the intensities of a set of weakly coupled harmonic oscillators are equal to those of an isolated, uncoupled set of equal density. The computed values are in the range of 2ϵ as expected for oxide ion vibrations.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Influence of Exchange and Correlation on Electron Transport in Metals*

F. J. Blatt
Physics Department, University of Illinois, Urbana, Illinois
(Received May 6, 1955)

A result of the collective treatment of Bohm and Pines is that the dependence of electron energy ϵ on the wave vector k differs from that for noninteracting free electrons. It follows that the expressions for those properties, such as conductivity and thermoelectric power, which depend on the relationship between ϵ and k must be suitably modified. Electrical and thermal conductivity are altered in the same manner, suffering slight changes because of changes of the density of states at the Fermi energy and of the relaxation time. The Wiedemann-Franz ratio remains unaltered. For the alkali metals the calculated thermoelectric power is reduced slightly below the free-electron value. The changes in both conductivity and thermoelectricity are too small to permit quantitative comparison with experiment.

I. INTRODUCTION

In a series of papers[1] Bohm and Pines have developed a new method for treating exchange and correlation effects in an electron gas. It was found that the influence of exchange and long-range correlation on the one-electron energies results in an energy *versus* wave vector relationship which is somewhat different from that given by the Hartree approximation. The collective description of electron interaction appears to be much superior to the Hartree-Fock approximation which neglects correlation of electrons with antiparallel spins and leads to incorrect results. In particular, near the Fermi energy the following relation holds[2]

$$\epsilon(k)=\frac{3.68k^2}{r_s^2}(m/m^*)-\frac{0.611}{r_s}\times\left[1-2\beta+\frac{\beta^2+3k^2-1}{2k}+\frac{1-k^2}{k}\ln\left(\frac{1+k}{\beta}\right)\right], \quad (1)$$

where ϵ is in rydbergs, r_s is the average interelectronic distance in units of the Bohr radius, and k is the wave vector in units of k_F, k_F being the wave vector at the Fermi energy. β is a parameter in the theory of Bohm and Pines which is a measure of the screening of the Coulomb interaction. Its value is given by $\beta=0.353\sqrt{r_s}$.[3] Equation (1) is to be compared with the free-electron expression

$$\epsilon_0(k)=\frac{3.68k^2}{r_s^2}\left(\frac{m}{m^*}\right). \quad (2)$$

The purpose of this note is to consider what changes appear in the expressions for the transport properties of an electron gas when Eq. (1) is used in place of the more common Eq. (2).

II. CONDUCTIVITY

The electrical and thermal conductivities of a metal are proportional to the density of states at the Fermi energy, $N(\epsilon_F)$, and to the relaxation time, τ.

The dependence of the relaxation time on the relationship between ϵ and k is a function of the type of scattering (lattice or impurity) and of the temperature. For lattice scattering in the high-temperature limit, an examination of the standard calculation of relaxation time[4] shows that

$$\tau=Dk\left(\frac{d\epsilon}{dk}\right)^2\Big/\frac{d^2\epsilon}{dk^2}, \quad (3)$$

* Supported by Office of Naval Research.

[1] D. Bohm and D. Pines, Phys. Rev. **82**, 625 (1951); **85**, 836 (1952); **92**, 609 (1953); D. Pines, Phys. Rev. **92**, 626 (1953).

[2] The term involving $\beta^3/6$ which appears in D. Pines, Report to the Solvay Congress, Eq. (28), has been omitted in Eq. (1) because it is a higher-order term which is canceled by other higher-order corrections. The author would like to thank Dr. Pines for informing him of this result before publication.

[3] D. Pines, private communication.

[4] A. H. Wilson, *Theory of Metals* (Cambridge University Press, London, 1953), second edition, p. 263.

TABLE I. Conductivity and thermoelectric power of the alkali metals.

	Li	Na	K	Rb	Cs
r_s	3.22	3.96	4.87	5.18	5.57
$(m^*/m)^a$	1.45	0.98	0.93	0.89	0.83
β	0.634	0.703	0.78	0.80	0.834
$N(\epsilon_F)/N_0(\epsilon_F)$	0.837	0.90	0.935	0.945	0.965
τ/τ_0	1.28	1.08	0.96	0.93	0.89
σ/σ_0	1.07	0.97	0.90	0.88	0.86
S/S_0	0.83	0.90	0.93	0.94	0.95

[a] Calculated by H. Brooks, quoted by D. Pines, Phys. Rev. **95**, 1090 (1954).

where D is a constant which depends on the atomic mass of the metal, the Debye temperature, the lattice parameter, and the strength of the electron-lattice interaction. Bardeen and Pines[5] have investigated the electron-lattice interaction in the Bohm-Pines collective description and have found that the effective matrix element for the electron-phonon interaction is identical, for wavelength of interest in conduction phenomena, with that found previously by Bardeen[6] using a self-consistent field method which neglects exchange and correlation effects. The constant D in Eq. (3) is, therefore, not altered when exchange and correlation are included.

Substitution of Eq. (2) into Eq. (3) leads to the well-known dependence of the relaxation time on $\epsilon^{\frac{1}{2}}$. From Eq. (1), one finds

$$(d\epsilon/dk)_{\epsilon_F} \equiv X_1 = 2A + B[\tfrac{1}{2}\beta^2 - 2 + 2\ln(2/\beta)] \quad (4)$$

and

$$(d^2\epsilon/dk^2)_{\epsilon_F} \equiv X_2 = 2A + B[3 - \beta^2 - 2\ln(2/\beta)], \quad (5)$$

where $A = (3.68/r_s^2)(m/m^*)$ and $B = 0.611/r_s$.

The fractional change in the density of states at the Fermi energy, $N(\epsilon_F)/N_0(\epsilon_F)$, is given by

$$N(\epsilon_F)/N_0(\epsilon_F) = 2A/X_1. \quad (6)$$

This ratio also gives the relative change in the electronic specific heat, C/C_0, and its values for the alkali metals are listed in the fourth row of Table I.

The ratio of the calculated relaxation time to the free-electron value, τ/τ_0, was obtained for each alkali metal by substitution of the parameters in the first three rows of Table I into Eqs. (4), (5), and (3). These ratios are listed in the fifth row of Table I, and in the sixth appear σ/σ_0, the ratios of the calculated electrical conductivities to the free-electron conductivities in the high-temperature limit. Although the changes are significant, in the sense that they could easily be detected, there are, unfortunately, other uncertainties in the theory of electronic conduction in metals which prevent quantitative comparison with experiment.

The fractional change of the thermal conductivity due to exchange and correlation is identical to that of the electrical conductivity, since again only the change due to the relaxation time and the density of states is involved. It follows, therefore, that the Wiedemann-Franz ratio is not affected by these considerations. (A statement to the contrary by Kohler[7] is in error.)

III. THERMOELECTRIC POWER

The absolute thermoelectric power of a metal is given by

$$S = \frac{\pi^2 k_0^2 T}{3e}\left(\frac{d(\ln N\tau k^2)}{d\epsilon}\right)_{\epsilon_F}, \quad (7)$$

where e is the charge of the electron and k_0 is Boltzmann's constant. From Eqs. (2), (3), and (7) one readily obtains the well-known result

$$S_0 = \pi^2 k_0^2 T/e\epsilon_F, \quad (8)$$

since in the free-electron approximation the logarithmic derivative is just $3/\epsilon_F$.

It can be shown by a bit of manipulation that when Eq. (1) is used

$$\left(\frac{d(\ln N\tau k^2)}{d\epsilon}\right)_{\epsilon_F} = (5/X_1) + (X_2/X_1^2) + Bf(\beta)/X_1X_2, \quad (9)$$

where $f(\beta) = 3\beta^2 - 8 + 6\ln(2/\beta)$.

The last row of Table I lists S/S_0 for the alkali metals. The ratio S/S_0 is of the same order of magnitude as σ/σ_0, and the comment regarding quantitative comparison with experiment applies here all the more. The present theory of the thermoelectric effects in metals is unable to account for many of the phenomena even qualitatively. Thus, for example, the absolute thermoelectric power of Li and Cu are positive although the free-electron model predicts the opposite sign. From the rather small influence of exchange and correlation on thermoelectricity it follows that this refinement in the theory does not provide a mechanism for explaining these anomalies.

[5] J. Bardeen and D. Pines (to be published).
[6] J. Bardeen, Phys. Rev. **52**, 688 (1937).
[7] M. Kohler, Abhandl. braunschweig. wiss. Ges. **5**, 48 (1953).
[8] See reference 4, p. 204.

Precision Determination of Lattice Constants with a Geiger-Counter X-Ray Diffractometer*

A. SMAKULA AND J. KALNAJS
Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received June 3, 1955)

It is shown that the Geiger-counter x-ray diffractometer can be used for precision determination of lattice constants. The zero point of the goniometer (determined by two methods) and the position of the diffraction lines (2θ) were obtained with an accuracy of $\pm 0.001°$. The temperature was kept constant by a special unit within 0.05°C. An investigation of the systematic errors was made, which shows that the extrapolation of lattice constants *vs* $\cos^2\theta$, according to Wilson's equation, is permissible and the vertical divergence error, as computed by Eastabrook, lies within our tolerances for 2θ up to 165° and may be neglected. The lattice constants of Al, Ag, Ge, Si, CaF_2, CsI, TlCl, and TlBr, determined by extrapolation, are in good agreement with published data. The limiting factor in the accuracy of the lattice constant determination is not in the error of the diffractometer, but in the x-ray wavelength.

INTRODUCTION

FOR the precision determination of lattice constants, the following methods are in use: Straumanis' asymmetrical method,[1] the Kossel-van Bergen compensation method,[2] and symmetrical focusing method.[3] Lattice constants are obtained by these with a precision of a few units in the fifth decimal place. The discrepancy between the published data, however, is frequently ten times higher. The spectrometer method should give the same precision,[4] provided the zero position of the goniometer is located with a proper accuracy and a precise extrapolation of the lattice constant to $2\theta = 180°$ can be made. This has not been possible thus far, and the spectrometer method was therefore not recommended for high precision.

Preliminary results with the Geiger-counter x-ray diffractometer[5] encouraged us to use the spectrometer method. Our experience with this instrument and the lattice constants of eight cubic crystals are reported in this paper.

GEIGER-COUNTER X-RAY DIFFRACTOMETER

The instrument itself has been described extensively by Parrish and Hamacher[6,7] and recently by Klug and Alexander.[8]

The accuracy of the instrument depends to a high degree on its alignment.[7] Of particular importance is an accurate determination of the zero position of the receiving slit. For this purpose a zero-setting slit (Fig. 1), as recommended by the manufacturer,[7] was placed in the center of the goniometer axis and the intensity distribution determined by counting in steps of 0.01° of 2θ (Fig. 2). Using the same intensity on the right and the left side of the maximum, the center line was determined. The intersection of the center line with the intensity curve gives one zero position. In order to eliminate the eccentricity of the setting slit, the same measurement was repeated, turning the slit 180° and the second zero position determined. The mean value of both gives the true zero position within 0.001°.

This method has a disadvantage, however, in that the axis of the sample holder has to be exchanged for a special axis; thus an unknown error might be introduced. As an alternative method we used therefore a shielding wedge,[9] placed directly on the sample-holder axis (Fig. 3). Rotating this wedge from +0.05° to −0.05° in steps of 0.01°, the x-ray intensity was determined by counting for two wedge positions differing by 180° (Fig. 4). The intersection of the two curves gives the zero point position. We obtained again an accuracy of 0.001° of 2θ in good agreement with the previous method. The result of three determinations by each method are given in the Table I.

The zero position of the sample holder (2:1 setting)

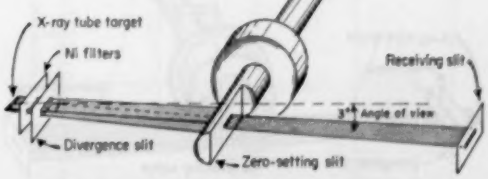


FIG. 1. Zero-setting slit for determination of the zero point of the goniometer.

* Sponsored by the U. S. Office of Naval Research, the U. S. Army Signal Corps, the U. S. Air Force, and the U. S. Ordnance Material Research Office.

[1] M. E. Straumanis, J. Appl. Phys. **20**, 726 (1949); see also: M. Straumanis and A. Ievins, *Die Präzisionsbestimmung von Gitterkonstanten nach der asymmetrischen Methode* (Springer-Verlag Berlin, 1940).

[2] H. van Bergen, Ann. Physik [5] **33**, 737 (1938).

[3] E. R. Jette and F. Foote, J. Chem. Phys. **3**, 605 (1935).

[4] A. J. C. Wilson, J. Sci. Instr. **27**, 321 (1950).

[5] This instrument is built by North American Philips Company and was called a "spectrometer" until 1952. A similar instrument is made by General Electric Company and recently by some European companies.

[6] W. Parrish and E. A. Hamacher, Transactions of Instruments and Measurements Conference, Stockholm, 1952 (unpublished).

[7] E. A. Hamacher and W. Parrish, Technical Report No. 42, Philips Laboratory, Irvington-on-Hudson, New York, April 26, 1951 (unpublished).

[8] H. P. Klug and L. E. Alexander, *X-Ray Diffraction Procedures* (John Wiley and Sons, Inc., New York, 1954).

[9] M. Tournarie, J. phys. et radium **15**, 11A (1954).

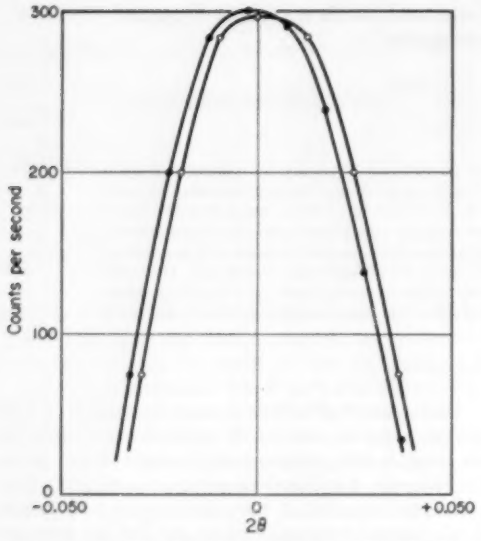


FIG. 2. Determination of the zero point of the goniometer by means of the zero-setting slit for two holder positions differing by 180°.

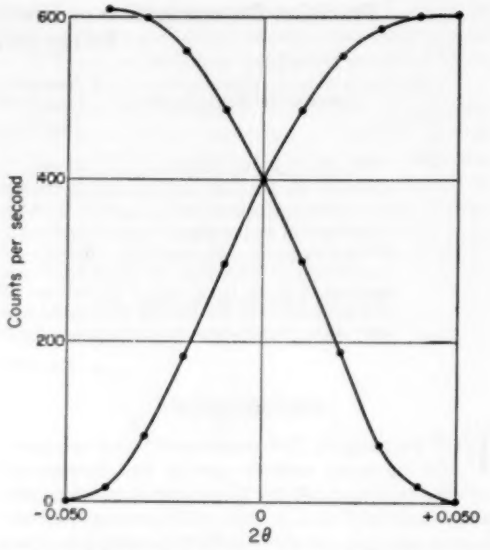


FIG. 4. Determination of the zero point of the goniometer by means of the shielding wedge for two holder positions differing by 180°.

was obtained by a special slit[7] (0.05 mm wide, 50 mm long) placed in the position of the sample holder (Fig. 5). An accuracy better than 0.01° was achieved, which is entirely adequate.

For lattice-constant precision higher than in the third decimal, the temperature of the sample, in general, must be controlled. The instrument as received from the factory does not have a temperature control, and therefore a special temperature chamber was built[10] (Fig. 6). This chamber connected to a Hoeppler ultrathermostat reduced temperature fluctuations to less than 0.05°C. The sample-holder axis was heated electrically to the same temperature when the room temperature was below 25°C in order to prevent a temperature gradient in the sample.

CRYSTALS AND SAMPLE PREPARATION

The materials investigated were: Al, Ge, Si, Ag, TlCl, TlBr, CaF_2, and CsI. This selection was made because Al is considered as the best x-ray standard material; Ge and Si are the standard semiconductors and can be made in very high purity; Ag, TlCl, and TlBr are strong absorbers for x-rays; and CaF_2 and CsI were not yet measured accurately. Only Ag was in microcrystalline form; the other materials were available in large single crystals.

The purity of the materials was: Ag[11] (99.999%) contained Cu, Pb, Fe, each <0.0001%, and Mg ≃0.0001%; Al[12] (99.99+%) contained 0.001% Si and a few other elements each <0.0001%; Ge[13] and Si[13] were characterized by the resistivity of 40 and 100 ohm-cm, respectively; TlCl, TlBr, and CsI were prepared in this laboratory; the thallium halides, according to spectroanalysis, showed only traces of a few common elements; a reciprocal contamination of anions was

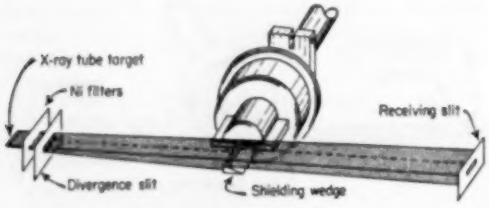


FIG. 3. Shielding wedge for determination of the zero point of the goniometer.

TABLE I. Accuracy of zero-point determination of goniometer by zero-setting-slit and shielding-wedge method.

	Zero-setting-slit 2θ	Shielding-wedge 2θ
	0.0075°	0.0065°
	0.0080	0.0067
	0.0073	0.0065
Average	0.0076°	0.0066°
Difference	0.001°	

[9] Smakula, Kalnajs, and Sils, J. Opt. Soc. Am. 43, 698 (1953).

[11] Material and spectroscopic analysis from Handy and Harman, Bridgeport, Connecticut.

[12] Material and spectroanalysis from Aluminum Company of America, New Kensington, Pennsylvania.

[13] Material and resistivity data from Raytheon Manufacturing Company, Waltham, Massachusetts (courtesy of Dr. H. Statz).

prevented by the use of Br-free HCl and Cl-free HBr for the synthesis; CaF_2[14] contained <0.001% of Ag, Fe, Mg, and Si, and somewhat more Sr; CsI was contaminated by Fe, K, Mg, Na, Sr, and V to <0.001% and by Rb about 0.01%.

The powder of the soft crystals (Al, TlCl, TlBr) was prepared by filing with a fine jeweler saw, using light strokes to prevent excessive heating. The harder crystals (Si, Ge, CaF_2, CsI) were ground in an agate mortar. The powder of all crystals except Si and Ge was annealed for 15 to 20 hours (Al, Ag, and CaF_2 at 500°C and CsI, TlCl, and TlBr at 300°C) in sealed Pyrex tubes, then sieved through a 325-mesh screen and packed gently in a Pyrex specimen holder with flat polished surfaces on an optically flat support. A few drops of a dilute binder (Lucite, collodion, or oil) were added when necessary. The powder specimens of TlCl and TlBr, not requiring a binder, were baked over night at *ca* 200°C with the sample top on an optical flat to prevent surface deformation. The packing density of powders was *ca* 65 percent.

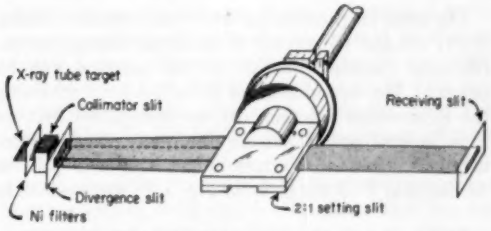


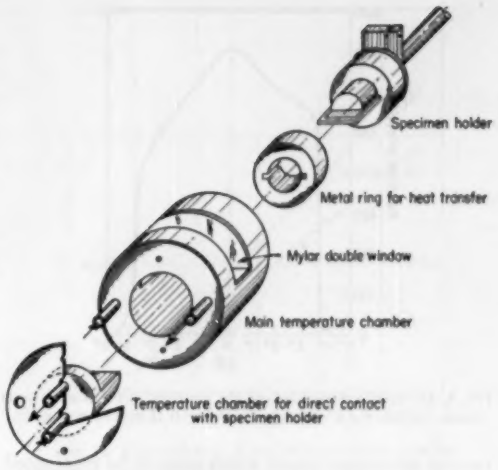
FIG. 5. Determination of the zero positions of the sample by means of (2:1) setting slit.

LOCATION OF DIFFRACTION ANGLES

A preliminary recording of the diffraction spectrum for each crystal gave the approximate location and intensity of the lines. After temperature equilibrium had been reached and the equipment stabilized, final measurements were taken by counting, scanning in steps of 0.025° or 0.05° in 2θ. The step setting was done by motor drive (¼° per minute) from high to low angles. This method gave a better reproducibility than setting by hand. The total number of counts was always greater than 10 000 for each point. This corresponds to an average error of about 0.67 percent. The counting rate was not greater than 600 counts per second, that is, in the linear range of the Geiger counter.

The peaks of the diffraction lines, which were used only for lattice-constant determinations, are found graphically by extrapolating the center lines to the maximum of the intensity curves (Fig. 7). The accuracy of 2θ determined this way is ±0.001°. Only sharp, single, and well-resolved lines in the back reflection region $2\theta > 120°$ were used. The zero point of the goniometer was rechecked frequently and the aperture

[14] From Optovac, North Brookfield, Massachusetts.


Specimen holder
Metal ring for heat transfer
Mylar double window
Main temperature chamber
Temperature chamber for direct contact with specimen holder

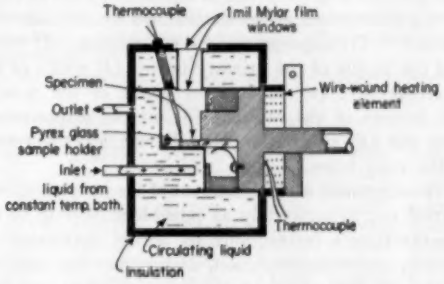


FIG. 6. Constant temperature chamber.

(4°) and the receiving slit (0.006 in.) not touched during the measurements. The short-anode Cu tube was operated with 35 kvp and 15 ma and nickel filter.

SYSTEMATIC ERRORS

Although the diffraction angles can be determined very accurately, the computed lattice constants contain

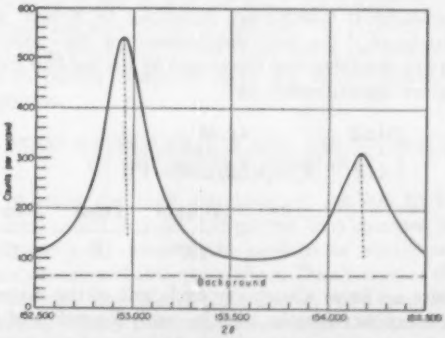


FIG. 7. Graphical determination of the peak of diffraction-line profile (711/551) for Cu $K\alpha_1$ and Cu $K\alpha_2$ for germanium.

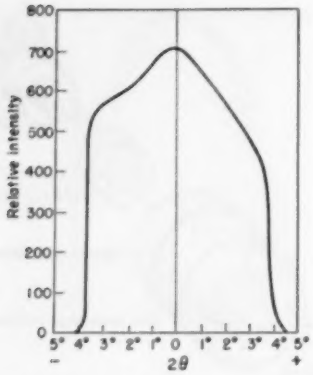


FIG. 8. Intensity distribution of the primary beam from a short anode Cu tube for 4° divergence slit and 0.15 mm receiving slit.

various systematic errors which have to be eliminated. The position and the shape of diffraction lines as given by the diffractometer are determined by the following factors[4,15]: (1) alignment of the instrument; (2) width and the profile of the primary beam; (3) width of the receiving slit; (4) penetration depth of the x-rays; (5) flatness of the sample; (6) sample displacement from the axis of rotation; and (7) vertical divergence of the x-ray beam.

The alignment and adjustment of the instrument was carried out as accurately as possible according to the manufacturer's instructions to obtain maximum intensity, optimum resolution, and correct line position.

The effective width of the primary beam was 0.06 mm. The vertical intensity distribution of the primary beam of our x-ray tube is shown in Fig. 8. It proves to be unsymmetrical; the shoulder is steeper on the low angle side. This might have an influence on the asymmetry of the diffraction lines at high angles.

The receiving slit used was 0.15 mm, that is, 2.5 times the effective width of the x-ray source, but still only about 1/10 of the half-width of the diffraction lines.

While the first three errors cause only a broadening of the lines, the next four cause a displacement and asymmetrical broadening. According to Wilson[4] and Eastabrook,[15] the total displacement of the center of gravity of a diffraction line caused by the last four errors is given approximately by

$$\Delta\theta = -\left\{\frac{\sin 2\theta}{4\mu R} - \frac{t\cos\theta}{R[\exp(2\mu t/\sin\theta)-1]}\right\} - \frac{A^2\sin 2\theta}{6R^2} - \frac{s\cos\theta}{R} - \frac{\delta^2\cot 2\theta}{12},$$

where μ = linear absorption coefficient of the material for x-ray wavelengths used in cm^{-1}; R = radius of the goniometer in cm; t = critical thickness of the sample $= (3.2/\mu)(\rho/\rho')\sin\theta$ in cm; ρ = density of solid material in g/cm^3; ρ' = density of the powder in g/cm^3; A = one-half the effective length of the specimen in cm; δ = spacing of the Soller slits divided by their length; and s = displacement of the front face of the specimen from the axis of rotation in cm.

The bracketed member in the above equation corrects for the penetration depth; the next, for the flatness of the sample; the third, for the sample displacement from the axis of rotation; and the fourth, for the vertical divergence of the x-ray beam. The first three are always negative and cause the diffraction lines to shift toward smaller angles. The shifting decreases with increasing 2θ and becomes zero for $2\theta = 180°$. The behavior of the vertical divergence error is different; it is negative for $2\theta < 90°$, zero for $2\theta = 90°$, and positive for $2\theta > 90°$. In the back-reflection region of interest here, the errors 4, 5, and 6 oppose that of 7; hence a diffraction angle exists at which the sum of the error goes through zero.

COMPARISON OF THEORETICAL AND EXPERIMENTAL SYSTEMATIC ERRORS

The error for penetration depth was computed, using $R = 17$ cm and 65 percent of the linear absorption coefficients (packing density of the samples was 65 percent). The second member in the bracket, containing the exponential function, can be disregarded because it is several orders smaller than any of the other members. Computed errors of penetration depth for Al, Ge, and TlCl are given in Fig. 9. It was possible to

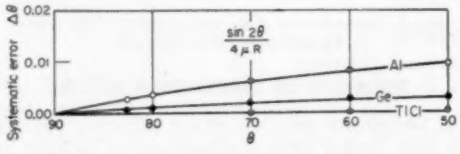


FIG. 9. Computed errors of penetration depth for Al, Ge, and TlCl.

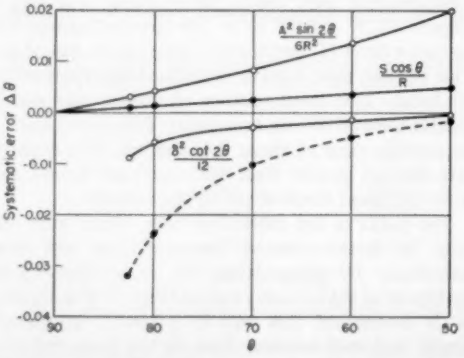


FIG. 10. Computed errors for flatness of the sample $A^2\sin 2\theta/6R^2$, for sample displacement from the axis of rotation $s\cos\theta/R$, and for vertical divergence $\delta^2\cot 2\theta/12$; dashes for $\delta = 0.04$; solid line for $\delta = 0.02$.

[15] J. N. Eastabrook, Brit. J. Appl. Phys. 3, 349 (1952).

check the error of penetration depth experimentally by comparing the computed and experimental differences for Al (low absorbing) and TlBr (high absorbing). The values agree within 20 percent.

The error caused by specimen flatness is computed for $A_0=0.6$ cm, using the change of A with 2θ according to $A=A_0/\cos(90-\theta)$. The effective length A of irradiated area is not constant, but varies with 2θ, having a minimum A_0 at $2\theta=180°$. An experimental check of this error was made on Ge with 1° and 4° apertures. A discrepancy by a factor three was found. This is, however, not surprising since the theoretical value of the error for 1° aperture cannot be correct. The derivation is valid only as long as the penetration depth is small compared with the effective length of the sample, which is not the case for 1° aperture.

The sample displacement from the center of rotation is somewhat uncertain. It is reasonable to assume $s=0.01$ to 0.03 mm.[4] Errors for specimen flatness and displacement from the center are given in Fig. 10.

In the vertical divergence error the constant δ was determined by measuring the total width of the collimator slit set (in this case 10 mm), dividing it by 18 spacings, and correcting for the thickness of the spacers, $d=0.025$ mm. We obtain for the single slit width $w=0.53$ mm. Since the length of the slit is 12.5 mm, $\delta=0.53/12.5=0.042$. The error computed with this value of δ is for $2\theta>140°$ higher than the sum of three others and increases rapidly with 2θ (Fig. 10). If this magnitude of the error is true, a strong decrease of the lattice constant *vs* 2θ would take place. This has not been observed (Fig. 11). The lattice constant continuously decreases with 2θ without any indication of an inflection. We assume therefore that the vertical divergence error in the region of $2\theta=160°$ is approximately equal to the sum of the three other errors or smaller; we set the effective value of δ to one half of the geometrical value.

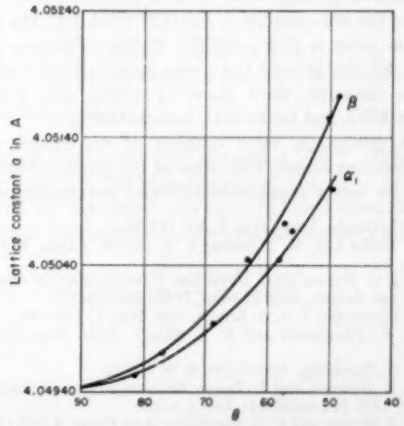


FIG. 11. Lattice constant of aluminum computed from various diffraction angles for Cu $K\alpha_1$ and Cu $K\beta$.

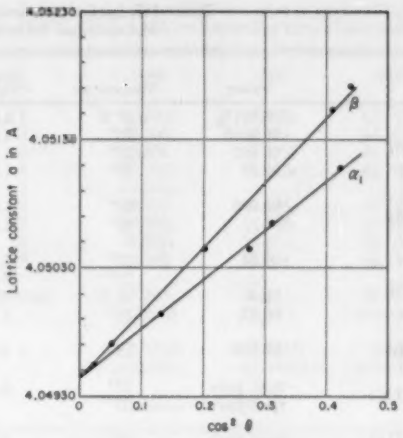


FIG. 12. Extrapolation of lattice constant *vs* $\cos^2\theta$ for aluminum for Cu $K\alpha_1$ and Cu $K\beta$.

TABLE II. Lattice constants determined by the Geiger-counter diffractometer using extrapolation and Cu $K\alpha_1(\lambda=1.54051\text{A})$ in angstroms at 25.0°C.

	a_{measured}	Correction for refractive index	a_{corr}	Δa error random	Δa error λ
Al	4.04957	0.00003	4.04960	0.00002	0.00016
Ag	4.08609	0.00012	4.08621	0.00002	0.00016
Si	5.43068	0.00004	5.43072	0.00002	0.00022
Ge	5.65745	0.00009	5.65754	0.00002	0.00023
TlCl	3.84240	0.00007	3.84247	0.00002	0.00015
TlBr	3.98580	0.00008	3.98588	0.00002	0.00016
CaF_2	5.46337	0.00005	5.46342	0.00002	0.00022
CsI	4.56766	0.00006	4.56772	0.00002	0.00018

The experimental determination of total errors was done in the following way. The lattice constants computed from various diffraction angles were plotted *versus* $\cos^2\theta$ according to Wilson.[4] This plotting gives straight lines as shown in Fig. 12 for Al. The extrapolation to $2\theta=180°$ gives the lattice constants a, corrected for systematic error. From lattice constants so obtained the diffraction angles were computed for the same Miller indices for which the diffraction lines were found with the diffractometer. The difference gives the total systematic errors. They are of the same order as the errors computed from the Wilson-Eastabrook equation.

FINAL DETERMINATION AND THE ACCURACY OF THE LATTICE CONSTANTS

In using the $\cos^2\theta$ extrapolation one has to keep several points in mind. Diffraction lines obtained with different x-ray wavelengths have to be plotted separately because of absorption effect. The absorption effect is shown in Figs. 11 and 12 for Cu $K\alpha_1$ and β lines. If the packing density from sample to sample is different, the position of the lines will vary. Points obtained with different apertures will not lie on the same line. All

TABLE III. Lattice constants from the literature recomputed for 25.0°C and corrected for new wavelengths according to Lonsdale.[a]

	Purity	Temperature	Original a in kx units	Error	Exp. coefficients	Recomputed a at 25°C
Al	99.971%	25°C	4.04139	±0.00008	...	4.04958[b]
	99.9986	25°	4.04145	±0.00002	23.29×10^{-6}	4.04963[c]
	99.992	25°	4.04134	...	...	4.04953[d]
	99.99	20°	4.04091	±0.00006	...	4.04958[e]
Ag	99.999	25°	4.07787	±0.00008	18.72×10^{-6}	4.08613[b]
	...	25°	4.07784	±0.00003	...	4.08610[f]
Si	99.84	25°	5.41982	±0.00034	4.15×10^{-6}	5.43078[b]
	...	23°	5.41975	±0.00005	...	5.43078[f]
	99.9	18.5°	5.41964	±0.00005	...	5.43075[g]
	99.97	25°	5.42002	±0.00003	...	5.43100[h]
Ge	99.999	25°	5.646235	±0.000002	5.92×10^{-6}	5.65758[h]
TlCl	spec. pure	25°	3.83459	±0.00004	54.57×10^{-6}	3.84236[f]
	spec. pure	25°	...	±0.00010	53.5×10^{-6}	3.84270[i]
TlBr	spec. pure	25°	3.97778	±0.00004	51.2×10^{-6}	3.98584[f]
CaF_2	j	25°	...	±0.0002	...	5.4626[k]
CsI	l	...	...	±0.0001	48.6×10^{-6}	4.5678[m]

[a] See reference 16.
[b] See reference 3.
[c] See reference 17.
[d] See reference 18.
[e] See reference 19.
[f] See reference 1.
[g] See reference 20.
[h] See reference 21.
[i] See reference 22.
[j] Spectroscopic analysis: As, B, Fe, Mg, Si, and Sr <0.001% each. Ag and Cu <0.0001% each.
[k] See reference 23.
[l] Na and K 0.01% each, Rb 0.03%.
[m] See reference 24.

parameters have to be kept constant for each extrapolation.

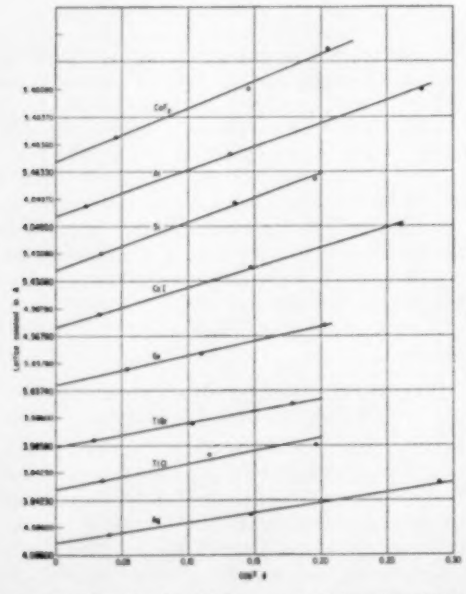


FIG. 13. Determination of lattice constants by extrapolation. A different ordinate is used for each crystal.

The extrapolation for the eight crystals are presented in Fig. 13 and the final lattice constants summarized in Table II. Column 2 shows the extrapolated values, column 3 corrections for refraction, and column 4 the corrected values. Table III gives the corresponding data from the literature.[1,3,16–24]

The accuracy of our lattice constants is determined by random and by systematic errors. The random errors are connected with the determination of diffraction angles. The average error of θ is $\pm0.001°$. Since the error in the zero position is $\pm0.001°$ (Table I), the total random error in θ is $\pm0.002°$. Using diffraction lines for $\theta>50°$ the $\Delta\theta$ error has a very small influence on the lattice constant since $\Delta a=-a\cot\theta\Delta\theta$. For $\theta=60°$, $\Delta a\simeq0.002\%$ and for $\theta=80°$, $\Delta a\simeq0.0006\%$.

The systematic error consists of wavelength and extrapolation errors. The value of Cu $K\alpha_1$ is 1.54051 A, based on newer measurements[25] and recommended by

[16] K. Lonsdale, Acta Cryst. **3**, 400 (1950).
[17] A. Ievins and M. Straumanis, Z. physik. Chem. **B34**, 402 (1936).
[18] A. J. C. Wilson, Proc. Phys. Soc. (London) **53**, 235 (1941).
[19] H. van Bergen, Ann. Physik [5] **39**, 553 (1941).
[20] H. Lipson and L. E. R. Rogers, Phil. Mag. [7] **35**, 544 (1944).
[21] M. E. Straumanis and E. Z. Aka, J. Appl. Phys. **23**, 330 (1952).
[22] P. G. Hambling, Acta Cryst. **6**, 98 (1953).
[23] H. E. Swanson and E. Tatge, National Bureau of Standards Report, 1953 (unpublished), Vol. 1, p. 69.
[24] T. B. Rymer and P. G. Hambling, Acta Cryst. **4**, 565 (1951).
[25] Y. Cauchois and H. Hulubei, *Tables des constantes et données numériques* (Hermann et Cie., Paris, 1947).

Lonsdale.[16] The wavelength error, according to Bragg,[26] is $\Delta\lambda = \pm 0.004\%$ ($\pm 0.001\%$ for kx and $\pm 0.003\%$ for the conversion factor λ_x/λ_A). Assuming the correctness of the extrapolation $vs\ \cos^2\theta$, the accuracy of the extrapolated values averages $\pm 0.0005\%$ or less. Thus the main error in the absolute lattice constants is caused by the uncertainty of x-ray wavelengths.

As an indication of wavelength uncertainty we show in Table IV the lattice spacings of Ge ($hkl = 711/511$) and of TlBr ($hkl = 510/431$) determined by Cu $K\alpha_1$ and Cu $K\alpha_2$, using the older λ values given by Bragg[27] and the newer by Lonsdale.[16] The new α_1 value is increased and α_2 decreased by 0.00001 A. Since α_1 and α_2 are very close, all instrumental errors will be practically the same. As one can see from Table IV, this small change in wavelength causes a change in spacing by 8 and 5 units in the fifth place. From the extrapolation lines, however, the difference in spacing should be not more than 2 units.

An objection can be made against the use of the peaks instead of the centers of gravity for determination of 2θ. According to Wilson, his formula is derived for the center of gravity, but at high diffraction angles it is valid for peaks also. In addition, as one can see from Fig. 8, the intensity distribution of the primary beam is unsymmetrical, contrary to the theory. Another influence on asymmetrical broadening is caused by vertical divergence which is also not included in Wilson's computation.

A comparison with published data obtained by Debye-Scherrer technique (see Table III) shows that our lattice constants are close to older data. However, a slight systematic discrepancy seems to exist. Our data are somewhat lower for low-absorbing crystals (Al, Si, Ge) and higher for high-absorbing (TlCl, TlBr, Ag). The deviation for high-absorbing crystals is probably caused by the absorption error in the Debye-Scherrer method. The reason for the opposite deviation in low-absorbing crystals is not known. A possibility of contamination by impurities was checked on Si and Ge. The data obtained for 9 Si and 7 Ge samples are shown in Tables V and VI. For all samples the average error in θ is less than 0.001° and the error in lattice constant ± 0.00001 A. Thus we can exclude a noticeable influence of impurities. The strongest support for the reliability of our lattice constants is given by good agreement of densities of our crystals computed from lattice constants and measured by weighing.[28]

We conclude that with the Geiger-counter diffractometer a precision in lattice constant measurement of ± 0.00002 A can be obtained, but that the absolute value is limited to 0.004 percent by the uncertainty of the x-ray wavelengths.

TABLE IV. Lattice constants of Ge (711/511) and TlBr (510/431), determined from Cu $K\alpha_1$ and α_2, using $\lambda_{\alpha_1} = 1.54050$ A and 1.54051 A, and $\lambda_{\alpha_2} = 1.54434$ A and 1.54433 A.

	λ according to Bragg[a]			λ according to Lonsdale[b]		
	$\lambda = 1.54050$	$\lambda = 1.54434$	Δa	$\lambda = 1.54051$	$\lambda = 1.54433$	Δa
Ge	5.65752	5.65751	0.00001	5.65756	5.65747	0.00009
TlBr	3.98583	3.98579	0.00004	3.98585	3.98576	0.00009

[a] See reference 27.
[b] See reference 16.

TABLE V. θ values for three diffraction lines of various Si crystals at 25.0°C, Cu $K\alpha_1$ and corresponding lattice constants.

No.	(444) θ°	(533) θ°	(620) θ°
1	79.303_8	68.432_6	63.757
2	79.306	68.434	63.755
3	79.305	68.432_5	63.759
4	79.305_5	68.432_6	63.760_2
5	79.306_5	...	...
6	79.307	68.434_2	63.757_8
7	79.308_2	68.430_8	63.755_8
8	79.307_5	68.431_5	63.756_2
9	79.308	68.429	63.754_5
Average	79.306_4	68.432_2	63.757
	± 0.0005	± 0.0006	± 0.0007
a	5.43080	5.43117	5.43135
	± 0.00001	± 0.00002	± 0.00003

$a_{extrap} = 5.43068$ A

TABLE VI. θ values for three diffraction lines of various Ge crystals at 25.0°C, Cu $K\alpha_1$ and corresponding lattice constants.

No.	(711/551)	(444)	(533)
1	76.477_8	70.600	63.216
2	76.478	70.597_8	63.214
3	76.478	70.601	63.219_5
4	76.479	70.601_5	63.217_5
5	76.477	70.602_5	63.217
6	76.479	...	...
7	76.477_5	70.601_5	63.217_5
Average	76.478_0	70.600_7	63.216_9
	± 0.0003	± 0.0007	± 0.0007
a	5.65755	5.65769	5.65788
	± 0.00001	± 0.00003	± 0.00003

$a_{extrap} = 5.65745$ A

[26] W. L. Bragg, J. Sci. Instr. **24**, 27 (1947).
[27] W. L. Bragg, Acta Cryst. **1**, 46 (1948).
[28] Smakula, Kalnajs, and Sils, Phys. Rev. **99**, 1747 (1955), this issue.

Precision Density Determination of Large Single Crystals by Hydrostatic Weighing*

A. SMAKULA AND V. SILS
Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received June 3, 1955)

In connection with a study of crystal imperfections, the density of Si, quartz, CaF_2, CsI, Ge, TlCl, TlBr, and polycrystalline Al have been determined by an hydrostatic weighing method using ethylene bromide as the immersion liquid. Random and systematic errors have been discussed and the present accuracy of the order of 10^{-5} reached.

INTRODUCTION

DENSITY combined with lattice constant has often been used for a study of crystal imperfections. Unfortunately, in most cases the two types of measurements have been carried out on different specimens or with insufficient accuracy. Therefore the results are not conclusive.

In our study of crystal imperfections, a group of large single crystals was selected. In the foregoing paper[1] were presented the lattice constant measurements; here we report our density determinations.

MATERIAL

The following single crystals (all cubic except quartz) have been used in our investigation: Si, quartz, CaF_2, CsI, Ge, TlCl, TlBr, and, in addition, a polycrystalline sample of Al.[2] Of these, Al is face-centered, CsI, TlCl, and TlBr are body-centered, Si and Ge have the diamond structure and CaF_2 has a structure similar to diamond. The densities vary considerably from 2.70 to 7.45 g/cm³. Also this group of crystals represents the three main bonding types: metallic, ionic, and covalent.

Si crystals[3] were grown in argon and Ge crystals[3] in nitrogen, both from the highest purity materials by the pulling method with a speed of 10 cm/hr. CaF_2 crystals[4] were grown from selected natural fluorite in a graphite crucible by the Stockbarger method in vacuum, annealed at 800°C for 24 hr and cooled down at a rate of 10°C/hr. CsI was prepared from CsBr, which was converted to $CsIO_3$ and this decomposed to CsI. The crystals of CsI were grown by the Bridgman method in sealed Vycor crucibles in an argon atmosphere with a speed of 1 mm/hr. TlCl and TlBr were prepared from pure Tl metal and purified by vacuum distillation. Crystals of TlCl and TlBr were grown in sealed Pyrex crucibles in an argon atmosphere with a speed of 1 mm/hr.

The strain effect on density was checked on thallium halides, which have the highest expansion coefficients. We were not abie to detect any density difference between samples annealed at 380°C for 24 hr and cooled down at a rate of 3°C/hr and unannealed samples. CsI was annealed immediately after growing for 12 hr at 500°C and cooled down at a rate of 15°C/hr. The vacuum-cast Al sample[5] was annealed in an evacuated Pyrex container at 480°C for 24 hr and cooled down at a rate 6°C/hr.

The purity of the crystals was: Al (99.99+%) contained 0.001% Si and a few other elements, each <0.0001%; Si and Ge were characterized by their resistivity of 100 and 40 ohm-cm, respectively; TlCl and TlBr, according to spectroanalysis, showed only traces of a few common elements, a reciprocal contamination by anions was prevented by the use of Br-free HCl and Cl-free HBr[6] for the synthesis; CsI was contaminated by Fe, K, Mg, Na, Sr, and V to <0.001% and by Rb about 0.01%; and CaF_2 contained <0.001% of Ag, Fe, Mg, and Si, and somewhat more Sr.

METHOD OF DENSITY DETERMINATION

Since our samples were large (1 to 20 cm³), the method of hydrostatic weighing was selected; an Ainsworth's semimicro balance, 100 g maximum load on each pan, rated sensitivity 0.01 mg, was used. The set of weights (Ainsworth, class *M*) was calibrated by the National Bureau of Standards. For volume determination the samples were suspended on a stainless steel wire (2 mils diameter) from the left pan and submerged in the thermostated liquid (Fig. 1). The bath temperature was measured with a thermometer (subdivisions 0.02°C) calibrated by the National Bureau of Standards. The temperature fluctuation in the bath was ±0.01°C and that of the immersion liquid ±0.005°C. Because of the water solubility of some crystals, ethylene bromide ($C_2H_4Br_2$) (b.p. 131.6° to 131.8°C at 760 mm Hg) served as immersion liquid. The density of ethylene bromide was determined by a secondary standard, a polished cube of natural quartz crystal. The density of

* Sponsored by the Office of Naval Research, the U. S. Army Signal Corps, the U. S. Air Force, and the U. S. Ordnance Material Research Office.

[1] A. Smakula and J. Kalnajs, Phys. Rev. **99**, 1737 (1955), preceding article.

[2] Aluminum has been included because the lattice constant measured by various authors shows the best agreement as compared with other materials.

[3] Material and resistivity data from Raytheon Manufacturing Company, Waltham, Massachusetts (courtesy of Dr. H. Statz).

[4] From Optovac, North Brookfield, Massachusetts.

[5] Material and spectroanalysis from Aluminum Company of America, New Kensington, Pennsylvania.

[6] We wish to thank Malinckrodt Chemical Works for the chlorine-free hydrobromic acid.

the quartz standard is referred to water as the primary standard.

DENSITY OF WATER

The standard water density is $d_{25°C}=0.997048$ g/cm³, as determined by Chappuis[7] and revised by Tilton and Taylor,[8] if 1.000027 ± 2[9] is used as the conversion factor for milliliter to cubic centimeters. The following factors influence the density of water: (1) The effect of dissolved air is noticeable in the 7th decimal place[10]; (2) the pressure effect is $\Delta d=+5\times10^{-5}$ per atmosphere[11]; for ordinary air-pressure fluctuation of ± 20 mm Hg, $\Delta d=\pm 1.3\times10^{-6}$; (3) the variation in the isotopic composition of natural water influences the density in the 7th place only,[12] provided the purification is done without isotopic fractionation; and (4) the influence of temperature is $\Delta d=2.5\times10^{-4}$ per °C at 25°C. A change of temperature of 0.01°C changes the density by 2.5×10^{-6}. Tap water was distilled once and

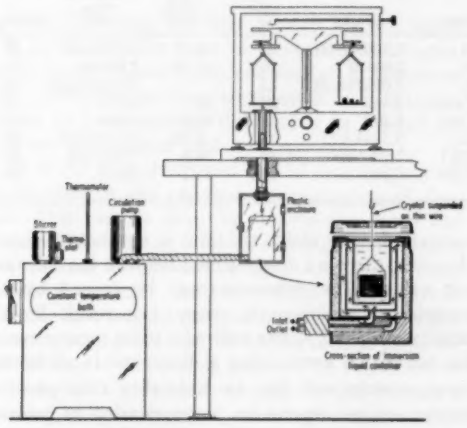


FIG. 1. Apparatus for density determination.

boiled before use; it had a conductivity of 1.6×10^{-6} (ohm-cm)$^{-1}$. According to Kirschenbaum[13] a change of 1×10^{-6} in conductivity causes a change of 1×10^{-6} in density.

DENSITY OF THE QUARTZ STANDARD

A polished cube (*ca* 16 cm³) of natural quartz crystal was used. After weighing in air the crystal was placed in a container (Fig. 1) with water and evacuated to remove air bubbles. For a good weighing reproducibility, the suspension wire was cleaned with a solution of potassium hydroxide in methyl alcohol. The height of water was kept at a constant level in all weighings. All weighings were corrected for air buoyancy. The humidity and temperature were measured inside the balance. The crystal density d has been computed applying all necessary corrections.[14] The mean value of five determinations is given in Table I and, for comparison, published data are listed also.[15–18]

TABLE I. Density of natural quartz crystals.

Density as measured	Temperature °C	Density computed for 25°C	Reference
2.650745	0	2.64847	15
2.64822±5	25	2.64822±5	16
2.65074	0	2.64846	17
2.64875±8	18	2.64810±9	18
2.64845±1.5	25	2.64845±1.5	present paper

The first value in Table I was obtained on large samples (*ca* 50 cm³) by comparing the weight with the mass standard (kilogram) and the volume with the length standard (meter). This value is considered the most reliable at present. The other three values are from pycnometric measurements on powdered material. For density computation for 25°C the following expansion coefficients[19] were used:

$$\alpha_{\parallel}=(7.161+0.0160\,t)\times10^{-6}/°C;$$
$$\alpha_{\perp}=(13.255+0.0223\,t)\times10^{-6}/°C.$$

DENSITY OF Si, Al, CaF_2, CsI, Ge, TlCl, AND TlBr

The density of the crystals was determined using ethylene bromide as the immersion liquid and quartz as the standard. To exclude any influence of room-temperature fluctuation on the density of ethylene bromide, the quartz standard was weighed before and after the sample weighing and the density of the liquid corrected, if necessary. The reproducibility of our meas-

TABLE II. Density of Si and Ge at 25°C.

Sample	Number of meas.	d_{H_2O} Si	Number of meas.	d_{H_2O} Ge
1	3	2.32901±0.00003	6	5.32673±0.00006
2	3	2.32900±0.00004	4	5.32674±0.00001
3	3	2.32903±0.00003	2	5.32672±0.00003
4	3	2.32904±0.00003		

[7] P. Chappuis, Trav. mém. bur. int. poid. mes. **14**, D1–D63 (DA) (1910).
[8] L. W. Tilton and J. K. Taylor, J. Research Natl. Bur. Standards **18**, 205 (1937).
[9] R. T. Birge, Revs. Modern Phys. **13**, 233 (1941).
[10] N. E. Dorsey, *Properties of Ordinary Water Substance* (Reinhold Publishing Corporation, New York, 1940).
[11] E. H. Amagat, Ann. chim. et phys. [6] **29**, 68, 505 (1893).
[12] E. W. Washburn and E. R. Smith, J. Research Natl. Bur. Standards **12**, 305 (1934).
[13] I. Kirschenbaum, *Physical Properties and Analysis of Heavy Water* (McGraw-Hill Book Company, Inc., New York, 1951), p. 295.
[14] See, e.g., N. Bauer in *Physical Methods of Organic Chemistry*, edited by A. Weissberger (Interscience Publishers, Inc., New York, 1949), second edition.
[15] I. Mace de Lepinay *et al.*, Trav. mém. bur. int. poid. mes. **14**, 120 (1910).
[16] P. H. Miller and J. W. M. DuMond, Phys. Rev. **57**, 198 (1940).
[17] T. Batuecas, Nature **165**, 62 (1950).
[18] T. Batuecas, Nature **173**, 345 (1954).
[19] I. R. Benoit, Trav. mém. bur. int. poid. mes. **6**, 1 (1888).

TABLE III. Densities of single crystals in g/cm³ determined by hydrostatic weighing. Standard: density of water at 25°C = 0.997048 g/cm³;[a] temperature = 25.00±0.01°C.

Crystal	Number of observations k	d	Errors random[b]	Errors systematic[c]	Errors total
Silicon	12	2.32902	1.0×15^{-6}	2.0×10^{-6}	3.0×10^{-6}
Quartz	5	2.64845	1.5	1.0	2.5
Aluminum (polycrystalline)	3	2.69801	1.5	1.5	3.0
CaF_2	3	3.17934	2.0	3.0	5.0
CsI	3	4.52593	8.0	4.0	12
Ge	4	5.32674	0.5	4.5	5.0
TlCl	3	7.01829	1.0	6.0	7.0
TlBr	3	7.45292	1.0	6.5	7.5

[a] The water density for 25.0°C as given by H. H. Landolt and R. Börnstein, *Zahlenwerte und Funktionen aus Physik Chemie Astronomie Geophysik und Technik* (Springer-Verlag, Berlin, 1955), Vol. 4, Part 1, p. 102, is $d_{25}=0.997046$ g/cm³. This small difference could cause a maximum change of 1.5×10^{-6} in densities which is within limits of our accuracy.

[b] Random error $=[\Sigma_i{}^k(d_k-d)^2/k(k-1)]^{\frac{1}{2}}$.

[c] Systematic error $=(d_{sample}/d_{quartz})\times$total error for quartz standard. The systematic error of the density of quartz was computed from the uncertainty of ±0.01°C in the temperature of the water.

urements for four silicon and three germanium samples from different crystals, and the random errors are given in Table II.

The measured densities are shown in Table III with random, systematic and total errors. Comparing the data for Si and Ge from Tables II and III, one finds that the difference between the various samples varies no more than the random error.

In Table IV are given the published densities of the crystals.[16,20–25] Only in the case of quartz is there an agreement between the older and our present values. All other values differ considerably. The difference may be caused by density variation of samples, influence of adsorbed air in powder measurements or temperature deviation.

CONCLUSIONS

Our measurements show that the density of large single crystals can be determined by hydrostatic weighing with an accuracy of 0.001 to 0.002 percent. Various specimens of silicon and germanium show constant density within the limit of our measurements. However, there is a disagreement between the older and our values. The difference may be caused by the material or experimental errors. Our result for the density of quartz agrees well with older measurements but not with a newer value of Batuecas. In all density measurements one has to remember that absolute density values depend on water density as primary standard, which is in error by at least 2×10^{-6}. It will be seen from the accompanying paper[26] that our weighing densities are in good agreement with densities computed from the lattice constants obtained from the same samples.

TABLE IV. Published densities.

Subst.	Density	Temperature °C	Lin. exp. coeff. $\times10^6$	Computed dens. 25°C	Reference
Si	2.32831±31	20	4.15	2.3282±3	20
Al	2.69892	20	23.29	2.69798	21
Al	2.69839±38	25	...	2.69839±38	16
CaF_2	3.180	20	19	3.179	22
CsI	4.509	25	48.6	4.509	23
Ge	5.3234±25	25	5.92	5.3234±25	20
TlCl	7.0000±4	30	54.6	7.0057±4	24
TlBr	7.540	21.7	51.2	7.537	25

[16] M. E. Straumanis and E. Z. Aka, J. Appl. Phys. **23**, 330 (1952).

[20] J. L. Snoek, Phil. Mag. **41**, 1188 (1950).

[21] H. E. Mervin, Am. J. Sci. [4] **32**, 429 (1911).

[22] G. P. Baxter and C. C. Wallace, J. Am. Chem. Soc. **38**, 264 (1916).

[23] E. Cohen and K. Piepenbrock, Z. physik. Chem. **A167**, 388 (1933).

[24] F. W. Clarke and H. Keck, J. Am. Chem. Soc. **5**, 241 (1884).

[25] Smakula, Kalnajs, and Sils, Phys. Rev. **99**, 1747 (1955), following article.

Densities and Imperfections of Single Crystals*

A. SMAKULA, J. KALNAJS, AND V. SILS
Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received June 3, 1955)

The densities of Si, Al, CaF_2, CsI, Ge, TlCl, TlBr, and SiO_2 (quartz) have been computed from lattice constants and molecular weights obtained from International Atomic Weights, and compared with the densities as determined by hydrostatic weighing of large single crystals. The hydrostatic density of Ge proves too large by 0.00193. This discrepancy disappears when the mass-spectroscopic atomic weight is used for the density calculation. Also for Si, Al, and CaF_2 a better agreement results with mass-spectroscopic atomic weights. Therefore, we conclude that the mass-spectroscopic atomic weights are more reliable than the officially accepted ones. From measured densities and lattice constants of Si, Al, CaF_2, and Ge, Avogadro's number is obtained as 6.02368×10^{23} mole^{-1} (chemical scale). The relative density defects for our samples of Si, Al, CaF_2, and Ge proves to be negligible within the limit of our measurements ($<5\times10^{-6}$). We consider these crystals the most perfect that we have obtained up to the present time. A somewhat higher relative density defect in quartz (11×10^{-6}) may be caused by uncertainty of the data used for the computation of the density and that for CsI (13×10^{-6}) by Rb contamination. TlCl and TlBr show the greatest relative density defects, 43×10^{-6} and 40×10^{-6}, respectively. These crystals may contain vacancies or dislocations in the order of $\sim5\times10^{13}$ per cc, as is known also for the silver halides.

INTRODUCTION

NATURAL and synthetic crystals contain imperfections in the form of impurities and crystal defects (lattice vacancies, dislocations, etc.). Since most of the physical properties are influenced by such imperfections to some extent, their study is one of the important problems of solid-state research.

One of the most direct methods of detecting crystal imperfections is the precision determination of densities. This can be done by determining mass and volume with a balance, or by calculating the mass from the molecular weight and Avogadro's number, using the lattice constant for the volume determination. While the first method gives an average of the density taken over a large volume, the second method gives the ideal density of the crystal lattice. The two densities can be equal only in ideal crystals. In real crystals a difference is to be expected, which gives information about the crystal imperfections. If the macroscopic density is higher than the microscopic, interstitials are present; if it is lower, vacancies (Schottky defects) or one or another kind of dislocations predominate. In case of impurities their effect on density can be positive or negative, depending on the ionic radius, atomic weight and how they are built into the lattice (interstitially or substitutionally). Instead of referring to densities, one can compare atomic weights with those obtained by chemical or mass-spectroscopic methods or compare Avogadro's number with a standard value.

Such investigations have been made in recent years by various authors.[1–5] The published results, however, differ considerably or are even contradictory. For instance, according to Straumanis[5] Al and Si[3] should have interstitials, but according to Addink[4] both should have vacancies. For Avogadro's number (chemical scale) Birge's[5] value is 6.02338×10^{23} mole^{-1}, Straumanis[5] uses 6.02403×10^{23} mole^{-1} and Addink's best value is 6.0228×10^{23} mole^{-1}. This study was therefore undertaken with the aim of deciding whether or not, at the present time, valid conclusions can be drawn on crystal imperfections from density determinations.

Our investigation differs from older ones in the following points:

(1) We used a group of crystals differing in lattice constants, densities and chemical bonding.

(2) The lattice constants and densities were measured on the same samples and at the same temperature.

TABLE I. Physical constants of crystals used for density computation.

	Crystal	n	$a_{18°C}$ [a]	Ref. lattice constant	Atomic or molecular weight: Internat.	Atomic or molecular weight: Mass-spec.[b]	Ref. mass-spec.
1	Si	8	5.43072 A	6	28.09	28.0875	10
2	Al	4	4.04960	6	26.98	26.98256	11
3	CaF_2	4	5.46342	6	78.08	78.079	12
4	CsI	1	4.56772	6	259.82	259.821	13, 14
5	Ge	8	5.65753	6	72.60	72.630	15
6	TlCl	1	3.84247	6	239.847	239.852	16
7	TlBr	1	3.98588	6	284.306	284.301	16
8	SiO_2	3	a: 4.91329	9	60.09	60.0875	10, 17
			c: 5.40483	9			

[a] For $CuK\alpha_1$ ($\lambda=1.54051$ A).
[b] Chemical scale.

* Sponsored by the Office of Naval Research, the U. S. Army Signal Corps, the U. S. Air Force, and the U. S. Ordnance Materials Research Office.

[1] Y. Tu, Phys. Rev. **40**, 662 (1932); D. A. Hutchison and H. L. Johnston, J. Am. Chem. Soc. **62**, 3165 (1940); Phys. Rev. **62**, 32 (1942); P. M. Miller and J. W. DuMond, Phys. Rev. **57**, 198 (1940); D. A. Hutchison, Phys. Rev. **66**, 144 (1944); J. Chem. Phys. **13**, 383 (1945); M. E. Straumanis, Z. Physik **126**, 49 (1949); Am. Mineralogist **38**, 662 (1953); Phys. Rev. **95**, 566 (1954); T. Batuecas, Nature **165**, 61 (1940); **173**, 345 (1954).

[3] M. E. Straumanis, Acta Cryst. **2**, 82 (1949).

[5] M. E. Straumanis and E. Z. Aka, J. Appl. Phys. **23**, 330 (1952).

[4] N. W. H. Addink, Rec. trav. chim. **70**, 202 (1951).

[5] R. T. Birge, Am. J. Phys. **13**, 63 (1945).

TABLE II. Avogadro's number as given in the literature from 1945 to 1954.

Reference	$N_A \times 10^{-23}$ mole^{-1} Physical scale	Chemical scale	Conversion factor phys. to chem. scale
1. Birge[a] 1945	6.02506±23	6.02338±23	1.000279[b]
2. DuMond and Cohen[c] 1948	6.0251±4	6.0235±4	1.000266[b]
3. Straumanis[d] 1949	6.02567±30	6.02403±30	1.000272[b]
4. Kirchner[e] 1950	6.0252±15	6.0236±15	1.000266[b]
5. DuMond and Cohen[f] 1950	6.025438±107	6.023762±107	1.000278[17]
6. Bearden and Watts[g] 1951	6.02566±16	6.02402±17	1.000272[b]
7. Witmer[h] 1951	6.02601±35	6.02433±35	1.000278[17]
8. Stille[i] 1952	6.02533±15	6.02365±15	1.000279[b]
9. DuMond and Cohen[j] 1952	6.02544	6.02376	1.000278[17]
10. DuMond and Cohen[k] 1954	6.02472±36	6.02304±36	1.000278[17]
11. Brogren (quartz)[l] 1954	6.02540	6.02372	1.000278[17]
12. Brogren (calcite)[l] 1954	6.02495	6.02327	1.000278[17]

[a] See reference 5.
[b] Obtained as ratio of physical to chemical scale.
[c] See reference 22.
[d] See reference 2.
[e] See reference 23.
[f] See reference 24.
[g] See reference 25.
[h] See reference 26.
[i] See reference 27.
[j] See reference 20.
[k] See reference 21.
[l] See reference 9.

(3) We used molecular weights computed from masses and abundances of atomic isotopes.

MATERIAL

Large single crystals of Si, Ge, CaF_2, TlCl, TlBr, CsI, and SiO_2 (natural quartz) and polycrystalline Al of very good quality were selected. Most of them belong to the cubic system; Si and Ge have diamond structure, CaF_2 is diamond-like; TlCl, TlBr and CsI have a body-centered structure; Al is face-centered cubic and SiO_2 (quartz) hexagonal. These crystals represent the three types of bonding: covalent, ionic, and metallic. In consequence, we expect various types of imperfections, expressing themselves in a different influence on lattice constant and density. The Al sample (vacuum-cast) was polycrystalline, but is included here because the published lattice constants for Al show the best agreement of all measured materials.[6] The purity of the crystals is given in the preceding paper.[7]

X-RAY DENSITIES AND COMPARISON WITH DENSITIES OBTAINED BY WEIGHING

The densities were computed for cubic crystals as $d_x = nA/a^3N$ (for quartz $d_x = 2nA/3a^2cN$), where n is the number of molecules per unit cell, A the atomic or molecular weight, a the lattice constant, and N Avogadro's number. The value for n, A, and a are listed in Table I. The atomic weights are those recommended by the International Commission on Atomic Weights,[8] and the lattice constants are those of the accompanying paper[6] and of Brogren.[9] In addition, in Table I atomic and molecular weights[10–17] computed from masses and abundances of atomic isotopes are given for later use.

The absolute error of the lattice constant is higher than the random error (±0.00002 A) because of the uncertainty of the x-ray wavelength by ±0.00006 A.[18] Cu $K\alpha_1$ ($\lambda = 1.54051$ A) was used, as recommended by Lonsdale.[19] For Avogadro's number the published data vary considerably. Table II contains the data published during the past ten years.[2,5,9,20–27] Some of these

[6] A. Smakula and J. Kalnajs, this issue [Phys. Rev. 99, 1737 (1955)].
[7] A. Smakula and V. Sils, this issue [Phys. Rev. 99, 1744 (1955)].
[8] E. Wichers, J. Am. Chem. Soc. 76, 2033 (1954).
[9] Computed for 25°C from data given for 18°C by G. Brogren, Arkiv Fysik 7, 47 (1954), using linear expansion coefficients

$$\alpha_{\parallel} = (7.161 + 0.0160\,t) \times 10^{-6}/°\text{C}$$

and

$$\alpha_{\perp} = (13.255 + 0.0223\,t) \times 10^{-6}/°\text{C}$$

taken from I. R. Benoit, Trav. mém. bur. int. poid. mes. 6, 1 (1888).

[10] Masses from: Duckworth, Preston, and Woodcock, Phys. Rev. 79, 188, 402 (1950). Abundances from: (a) M. G. Inghram, Phys. Rev. 70, 653 (1946); (b) E. P. Ney and J. H. McQueen, Phys. Rev. 69, 41 (1946); (c) D. Williams and P. Yuster, Phys. Rev. 69, 556 (1946); (d) J. R. White and A. E. Cameron, Phys. Rev. 74, 991 (1948); (e) R. F. Hibbs as reported by K. T. Bainbridge and A. O. Nier, Preliminary Report No. 9, Nuclear Science Series (National Research Council, Washington, 1950); (f) J. H. Reynolds, Phys. Rev. 90, 1047 (1953); average taken from (a) to (f).
[11] C. W. Li, Phys. Rev. 88, 1038 (1952).
[12] Masses for calcium from: Collins, Nier, and Johnson, Jr., Phys. Rev. 84, 717 (1951). Abundances for calcium from: J. R. White and A. E. Cameron, Phys. Rev. 74, 991 (1948).
[13] H. E. Duckworth, Nature 174, 595 (1954), Cs.
[14] R. E. Halsted, Phys. Rev. 88, 666 (1952), I.
[15] Masses from: Collins, Johnson, and Nier, Phys. Rev. 94, 398 (1954). Abundances from: (a) R. F. Hibbs as reported by K. T. Bainbridge and A. O. Nier, Preliminary Report No. 9, Nuclear Science Series (National Research Council, Washington, 1950); (b) Inghram, Hayden, and Hess, Report, Argonne National Laboratory Report, ANL-4082, 1947 (unpublished, p. 5); average taken of both. Because of large errors the recent data of J. H. Reynolds, Phys. Rev. 90, 1047 (1953) have not been used.
[16] T. L. Collins (private communication).
[17] A. O. Nier, Phys. Rev. 77, 789 (1950).
[18] W. L. Bragg, J. Sci. Instr. 24, 27 (1947).
[19] K. Lonsdale, Acta Cryst. 3, 400 (1950).
[20] J. W. M. DuMond and E. R. Cohen, Am. Scientist 40, 447 (1952).
[21] J. W. M. DuMond and E. R. Cohen, Revs. Modern Phys. 25, 691 (1953).
[22] J. W. M. DuMond and E. R. Cohen, Revs. Modern Phys. 20, 82 (1948).
[23] F. Kirchner, *Landolt-Börnstein Tabellen* (Springer-Verlag, Berlin, 1950), sixth edition, Vol. I, Part 1.
[24] J. W. M. DuMond and E. R. Cohen, Phys. Rev. 82, 555 (1951).
[25] J. A. Bearden and H. M. Watts, Phys. Rev. 81, 73 (1951).
[26] E. E. Witmer, Phys. Rev. 81, 308 (1951).
[27] U. Stille, Physik. Bl. 9, 397 (1952).

values were computed from densities and lattice constants obtained on different samples: others were derived from different physical constants. We decided to use Birge's value since it is based on the same kind of measurements as ours, and served as a starting value for the computation of the universal physical constants.[20,21]

Our densities are given (Table III) under d_{x1} computed using International Atomic Weights and under d_w those determined by hydrostatic weighing.[7]

(a) Influence of Atomic Weight on Density

In Fig. 1, Δd_{x1} is plotted *vs* d_w. For most crystals the difference Δd_{x1} proves to be smaller than 0.0008, except for Ge. Initially we assumed that the large deviation for Ge was caused by interstitials.[23] However, doubt arose when no comparable deviation was found for Si. The only other possibility for the great deviation in Ge was a corresponding error in the atomic weight. The value of the atomic weight given to Si has been changed in recent years[20] by three units in the second decimal

TABLE III. Comparison of densities determined by weighing (d_w), and computed from lattice constants (Table I), using molecular weights from international atomic weights (d_{x1}) and from masses and abundances of isotopes (d_{x2}), the Avogadro number of Birge (6.022338×10^{23} mole^{-1}), and temperature 25.0°C. $\Delta d_{x1}=d_{x1}-d_w$, $\Delta d_{x2}=d_{x2}-d_w$.

	d_w	d_{x1}	Δd_{x1}	d_{x2}	Δd_{x2}	$\Delta d_{x2}:d_w$
Si	2.32902	2.32932	0.00030	2.32912	0.00010	0.000043
Al	2.69801	2.69790	−0.00011	2.69815	0.00014	0.000052
CaF_2	3.17934	3.17955	0.00021	2.17951	0.00017	0.000053
CsI	4.52593	4.52620	0.00027	4.52622	0.00029	0.000064
Ge	5.32673	5.32480	−0.00193	5.32700	0.00027	0.000051
TlCl	7.01829	7.01880	0.00051	7.01894	0.00065	0.000093
TlBr	7.45292	7.45372	0.00080	7.45359	0.00067	0.000090
Quartz	2.64845	2.64866	0.00021	2.64855	0.00010	0.000038

place because of new determinations by chemical and mass-spectroscopic methods. The last atomic weight determination of Ge was made by a chemical method.[20] No mass-spectroscopic atomic weight of Ge has been published. Through the courtesy of Professor K. T. Bainbridge and Dr. T. L. Collins, Harvard University, the atomic weight for Ge was obtained as calculated from mass-spectroscopic data. This value, together with those for the other crystals, is listed in Table I. With these mass-spectroscopic molecular weights, the densities for our crystals assume the value d_{x2} (Table III). Δd_{x2} gives the density difference between these computed (d_{x2}) and the measured values d_w. As seen from Table III and Fig. 1, the great density difference in Ge disappears and the differences for most of the other crystals are much reduced. A slight increase occurs for TlCl. We can conclude that the density differences for most of our crystals were caused by the uncertainty of the atomic weights.

[23] Smakula, Kalnajs, and Sils, Phys. Rev. **97**, 253 (1955).
[20] E. Wichers, J. Am. Chem. Soc. **74**, 2446 (1952).
[21] Hönigschmid, Wintersberger, and Wittner, Z. anorg. u. allgem. Chem. **225**, 81 (1935).

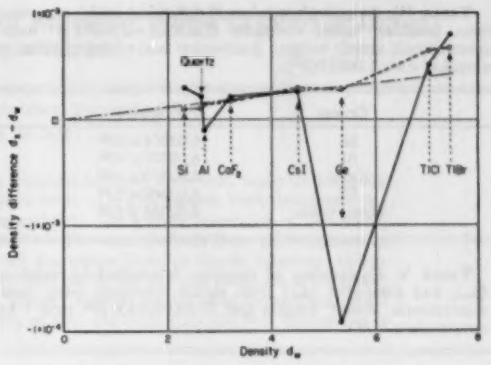


FIG. 1. Difference of x-ray and weighing density $d_x-d_w=\Delta d_x$. For computing x-ray density: Cu $K\alpha_1$ $\lambda=1.54051$ A and Birge's Avogadro number 6.02338×10^{23} mole^{-1}. Solid circles, Δd_{x1}, international atomic weights used; open circles, Δd_{x2}, mass-spectroscopic atomic weights used.

(b) Influence of Lattice Constant on Density

The random error of our lattice constants for all crystals is ±0.00002 A.[8] The systematic error caused by the uncertainty of the x-ray wavelength is 0.004 percent. Since we used the same wavelength (Cu $K\alpha_1$) in all our lattice-constant determinations, this error is a constant factor for all of the crystals. The lattice constants are influenced proportionally to their absolute values. The influence of the lattice-constant error on density can not be separated from the error of Avogadro's number.

(c) Influence of Avogadro's Number on Density

While A and a vary from crystal to crystal, N is a constant, therefore its influence on $\Delta d_{x2}:d_w$ must be a constant. In Fig. 2, $\Delta d_{x2}:d_w$ is plotted *vs* d_w. All values are positive and very nearly equal. This could be interpreted by vacancies. However, this interpretation would lead to the improbable consequence that all of our crystals (except TlCl and TlBr) have an equal number of vacancies. We therefore assume that our product a^3N is not correct. The absolute error of a and N is practically the same (∼0.0002). Since a is based on the crystallographic conversion factor $\lambda_{kx}:\lambda_{kx}$ which is at present generally accepted, we keep a constant

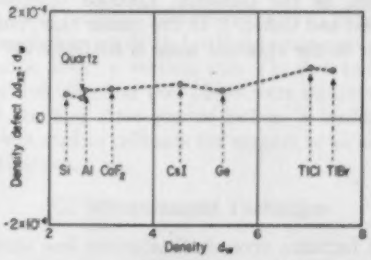


FIG. 2. Relative density defects ($\Delta d_{x2}:d_w$). (We used $N=6.02338\times10^{23}$ mole^{-1} and mass-spectroscopic atomic weights.)

TABLE IV. Avogadro's number N (chemical scale) computed from densities, lattice constants (Cu $K\alpha_1\lambda = 1.54051$ A) mass-spectroscopic atomic weights (conversion factor from physical to chemical scale 1.0002783[17]).

Crystal	N (mole^{-1})
Si	6.02364×10^{23}
Al	6.02370×10^{23}
CaF_2	6.02370×10^{23}
Ge	6.02369×10^{23}
Mean value:	6.02368×10^{23}

TABLE V. Comparison of densities determined by weighing (d_w), and computed (d_{x2}) from lattice constants using mass-spectroscopic atomic weights and $N = 6.02368\times10^{23}$ mole^{-1} for temperature 25.0°C.

	d_w	d_{x2}	Δd_{x2}	$\Delta d_{x2}:d_w$
Si	2.32902	2.32901	−0.00001	−0.0000043
Al	2.69801	2.69802	0.00001	0.0000037
CaF_2	3.17934	3.17935	0.00001	0.0000031
CsI	4.52593	4.52599	0.00006	0.000013
Ge	5.32673	5.32674	0.00001	0.0000019
TlCl	7.01829	7.01859	0.00030	0.000043
TlBr	7.45292	7.45322	0.00030	0.000040
Quartz	2.64845	2.64842	−0.00003	−0.000011

and correct N. N-values computed from four of our best crystals (Si, Al, CaF_2, and Ge) are given in Table IV. The mean value is $(6.02368\pm0.00002)\times10^{23}$ mole^{-1}.

Using this Avogadro's number, and atomic and molecular weights computed from mass-spectroscopic data and our lattice constants, the final densities d_{x2}, differences $\Delta d_{x2} = d_{x2} - d_w$ and $\Delta d_{x2}:d_w$ = relative density defects are computed (Table V and Fig. 3).

CONCLUSIONS

A study of crystal imperfections by density measurements requires a knowledge of atomic weights, lattice constants, and Avogadro's number to a high accuracy. We found that the international atomic weights are sometimes unreliable and have to be replaced by more accurate data obtained from masses and abundances of atomic isotopes. The values of lattice constants are based on the absolute values of the x-ray wavelengths; these are obtained from Siegbahn kX units by multiplication with a conversion factor $f = \lambda_s/\lambda_x$. We can use either the crystallographic conversion factor 1.002026, or the Universal 1.002063 (computed by DuMond and Cohen[21]). In the former case Avogadro's number in the chemical scale is 6.02368×10^{23} mole^{-1} and in the latter 6.02301×10^{23} mole^{-1}, which is very close to DuMond and Cohen's value 6.02304×10^{23} (see Table II). Only two of the previous N values, those from KCl[5] and from quartz, are close to ours; the value obtained from calcite by Brogren[9] is lower and that by Straumanis[2] is higher. The discrepancy for calcite is probably caused by impurities.[21,22] Calcite can hardly qualify as a standard crystal. Even if its impurities are determined, a correction for molecular weight cannot be made as long as it is not known whether impurities are bound substitutionally or interstitially.

Of the crystals studied here, Si, Al, CaF_2, and Ge contain relative density defects, $\Delta d_{x2}:d_w$, which are less than 5×10^{-6}, that is, undetectable within the limit of our measurements. The values 11×10^{-6} for quartz and 13×10^{-6} for CsI are still close to this limit. The only cases where we have an indication of imperfections are TlCl and TlBr. In both crystals the relative density defects are *ca* 4×10^{-5}. Relative density defects of 22×10^{-5} and 8×10^{-5} have been found recently in AgCl and AgBr single crystals.[22] Since silver and thallium halides are similar in various properties (broad ultraviolet absorption bands, high plasticity, low melting points, etc.), they may have the same kind of imperfections.

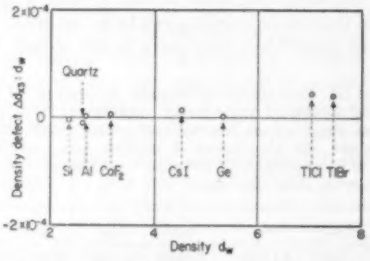


FIG. 3. Relative density defects ($\Delta d_{x2}:d_w$). (We used $N = 6.02368\times10^{23}$ mole^{-1} and mass-spectroscopic atomic weights.)

ACKNOWLEDGMENTS

We are greatly indebted to Professor A. R. von Hippel for his interest and advice in this work.

[21] A. Ievins and M. Straumanis, Z. Physik **116**, 194 (1940).
[22] K. W. Andrews, Mineralog. Mag. **29**, 85 (1950).
[23] C. R. Berry, Phys. Rev. **97**, 676 (1955).

Oscillatory Galvanomagnetic Properties of Antimony Single Crystals at Liquid Helium Temperatures

M. C. STEELE
United States Naval Research Laboratory, Washington, D. C.
(Received May 20, 1955)

Measurements of the Hall coefficient and transverse magnetoresistance have been made on antimony single crystals at liquid helium temperatures and in magnetic fields up to 25 kilogauss. Both these properties exhibit oscillations which are periodic in reciprocal magnetic field. For a given orientation of the crystal axes relative to the magnetic field, the period of oscillation is the same for both these galvanomagnetic effects. The observed periods are in good agreement with the values found in the de Haas-van Alphen effect of antimony. The magnetic field dependence of the Hall coefficient was also examined at 78°K and 300°K. It was found that at 78°K the Hall coefficient already exhibits marked field dependence. Small amounts of impurities were found to have a marked effect on both the magnitude and field dependence of the Hall coefficient at low temperatures.

INTRODUCTION

EXPERIMENTS have shown that the galvanomagnetic and thermomagnetic properties of some of the metals which exhibit the de Haas-van Alphen effect also show oscillatory behavior as a function of magnetic field, H. Prior to the present work the correlation for the period of oscillation in H^{-1} had been established for bismuth,[1–5] graphite,[6,7] and zinc.[6,8] These metals are characterized by relatively long periods of oscillation in so far as the de Haas-van Alphen effect is concerned. On the other hand, antimony, which also exhibits the de Haas-van Alphen effect, has a relatively short period of oscillation[9] (of the order of fifty times smaller than that for Bi, Zn, or graphite). In general the amplitude of the susceptibility oscillations decreases as the period of oscillations decreases, so that it was of interest to see if the corresponding oscillatory galvanomagnetic effects could be observed in antimony. This paper concerns itself primarily with the measurements of the magnetic field dependence of the Hall coefficient and the transverse magnetoresistance of Sb single crystals in fields up to 25 kilogauss at liquid helium temperatures. A preliminary report[10] of the discovery of the magnetoresistance and Hall effect oscillations has already been given.

EXPERIMENTAL DETAILS

A. Crystals

The antimony single crystals used for most of these experiments (Sb I–Sb IV) were cleaved from a crystal kindly donated and grown by W. Zimmerman, III of the Crystal Branch, U. S. Naval Research Laboratory. The original material, as obtained from the Bradley Mining Company, San Francisco, California, had a purity of 99.997 percent. Zone refinement made the purity of the final crystal better than 99.999 percent. Cleaving the crystal under liquid nitrogen readily located the trigonal axis. Then the series of plates of about 0.5 mm thickness were cleaved from the large crystal. For each of these plates, the trigonal axis was perpendicular to the face of the plate. Examination of these crystals showed that some twinning had resulted since three sets of lines making angles of 60° with each other were barely visible. These lines allowed the binary axis to be located relative to the long dimension of the experimental rectangular specimens cut from the plates by an air stream containing powdered abrasive. The length of each rectangle was .10 mm, while the width was 2 mm. Thickness of individual specimens will be given when the results are presented. The long dimension was mounted vertically in all the experiments.

One of the specimens (Sb V) used for these experiments was prepared in the above manner from a crystal kindly donated by Dr. E. I. Salkovitz, Metal Physics Branch, Naval Research Laboratory. The original material for this crystal was Johnson-Matthey antimony (Lab. No. 4470) which had a purity of 99.92 percent (by difference).

B. Magnet

Horizontal magnetic field strengths up to 25 kilogauss were supplied by an A. D. Little electromagnet capable of rotation about a vertical axis. The five and three-quarter inch diameter pole pieces were separated by a gap of one and three-quarter inches. A nuclear fluxmeter was used to calibrate the magnet to an accuracy of 0.02 percent.

C. Measurement Technique

Current and potential leads were attached to each specimen with low melting-point solder. The measuring current was always parallel to the long dimension of

[1] P. B. Alers and R. T. Webber, Phys. Rev. **91**, 1060 (1953).
[2] T. G. Berlincourt, Phys. Rev. **91**, 1277 (1953).
[3] Reynolds, Leinhardt, and Hemstreet, Phys. Rev. **93**, 247 (1954). See also L. C. Brodie, Phys. Rev. **93**, 935 (1954).
[4] M. C. Steele and J. Babiskin, Phys. Rev. **94**, 1394 (1954).
[5] J. Babiskin and M. C. Steele, Phys. Rev. **96**, 822 (1954).
[6] T. G. Berlincourt and J. K. Logan, Phys. Rev. **93**, 348 (1954).
[7] T. G. Berlincourt and M. C. Steele (to be published).
[8] N. M. Nachimovich, J. Phys. U.S.S.R. **6**, 111 (1942).
[9] D. Shoenberg, Proc. Roy. Soc. (London) **245**, 1 (1952).
[10] M. C. Steele, Phys. Rev. **98**, 1180(A) (1955).

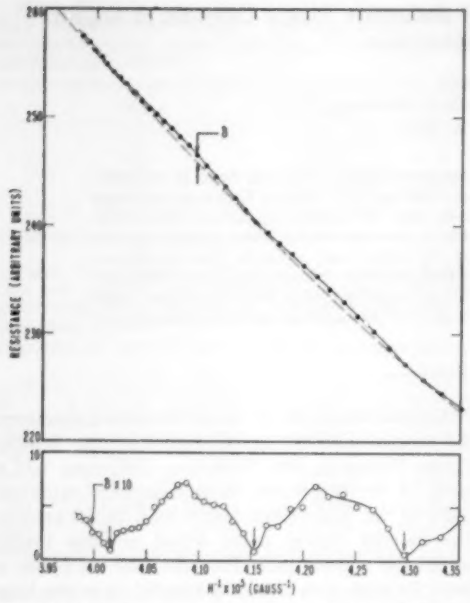


FIG. 1. Magnetoresistance of Sb I as a function of H^{-1} at 1.52°K. H was perpendicular to the trigonal axis and made an angle of 25° with a binary axis.

the specimens. The two potential probes used to measure resistance were on the same edge of the crystal, separated by a distance of about 3 mm. A third potential probe was located on the opposite edge to measure Hall voltages. Except for one experiment, Hall effect measurements were made only with H parallel to the trigonal axis. This geometry was ascertained by the visual location of the trigonal axis after the crystal was mounted. Attempts to locate the trigonal axis from magnetoresistance measurements proved to be an interesting problem in itself since the results depended very strongly upon the angle between a binary axis and the electric current. This had already been observed by Rausch[11] in his work on Sb single crystals down to liquid nitrogen temperatures. Our results support his conclusions and extend the rotation measurements to 4.21°K.

The Hall coefficient, A, is defined by the equation:

$$A = 10^8 Vt/IH, \tag{1}$$

where A is in cm³/coulomb; H is in gauss; the measuring current, I, is in amperes; the Hall potential, V, is in volts, and the crystal thickness, t, is in centimeters. Throughout these experiments the measuring current was either 200 or 400 milliamperes. Typical Hall voltages ranged from about 50 to 15 microvolts as the magnetic field was decreased from 25 to 6 kilogauss. These voltages were measured to an accuracy of 0.05 microvolt with a six-dial potentiometer in conjunction with a photoelectric galvanometer. Magnetoresistive components of the transverse voltages were separated from the Hall components by taking two sets of measurements in which the direction of H is changed by 180°. If we denote by $V(H)$ and $V(-H)$ the voltages measured in this way, then it is assumed that

$$2V_\rho = V(H) + V(-H), \tag{2}$$

$$2V_A = V(H) - V(-H), \tag{3}$$

where V_ρ is the magnetoresistance voltage component and V_A is the Hall voltage. This resolution supposes that the magnetoresistance is an even function of H. Measurements from the resistance probes confirmed this supposition over the range of field and temperature used. From Eqs. (2) and (3), simultaneous values of Hall coefficient and magnetoresistance are obtained. This is useful for comparing the relative values of the period and phase of the oscillations in these properties.

EXPERIMENTAL RESULTS

(A) Sb I

The first experiments with these crystals were restricted to a measurement of the magnetoresistance (from the resistance probes) as a function of H. Figure 1 shows the results at $T = 1.52$°K for Sb I plotted as a function of H^{-1}. This crystal had $t = 0.58$ mm. The angle, ψ, between the specimen length and a binary axis was 5°. H was perpendicular to the trigonal axis and made an angle of 25° with a binary axis. In order to emphasize the oscillatory nature of the resistance, the difference δ between the smooth dashed envelope curve and the solid curve connecting the experimental points is also plotted in Fig. 1. Our subsequent analysis will show that the δ curve is periodic in H^{-1}. Over the range of field shown in Fig. 1 the maximum value of the amplitude of the oscillating component ($\delta/2$) is only 0.15 percent of the total resistance. It is evident from the structure in the δ curve of Fig. 1 that there may be more than one oscillating term contributing to the phenomenon at this particular orientation.

(B) Sb IV

For this specimen t was 0.44 mm and ψ was 15°. The Hall coefficient, A, at $T = 1.57$°K is plotted as a function of H^{-1} in Fig. 2. H was parallel to the trigonal axis (to within 3°). The oscillations in A are very distinct; however, the modulation in the amplitude of the oscillations suggests that there is more than one oscillating term present. At 25 kilogauss, the amplitude of the oscillatory component of A is 0.6 percent of the mean value of A. Figure 3 shows the magnetoresistance as a function of H^{-1} under the same conditions as for Fig. 2. In fact, the data in Fig. 3 are obtained from the Hall effect measurements through Eq. (2). Again the oscillating component is easily discernible from the

[11] K. Rausch, Ann. Physik **1**, 190 (1947).

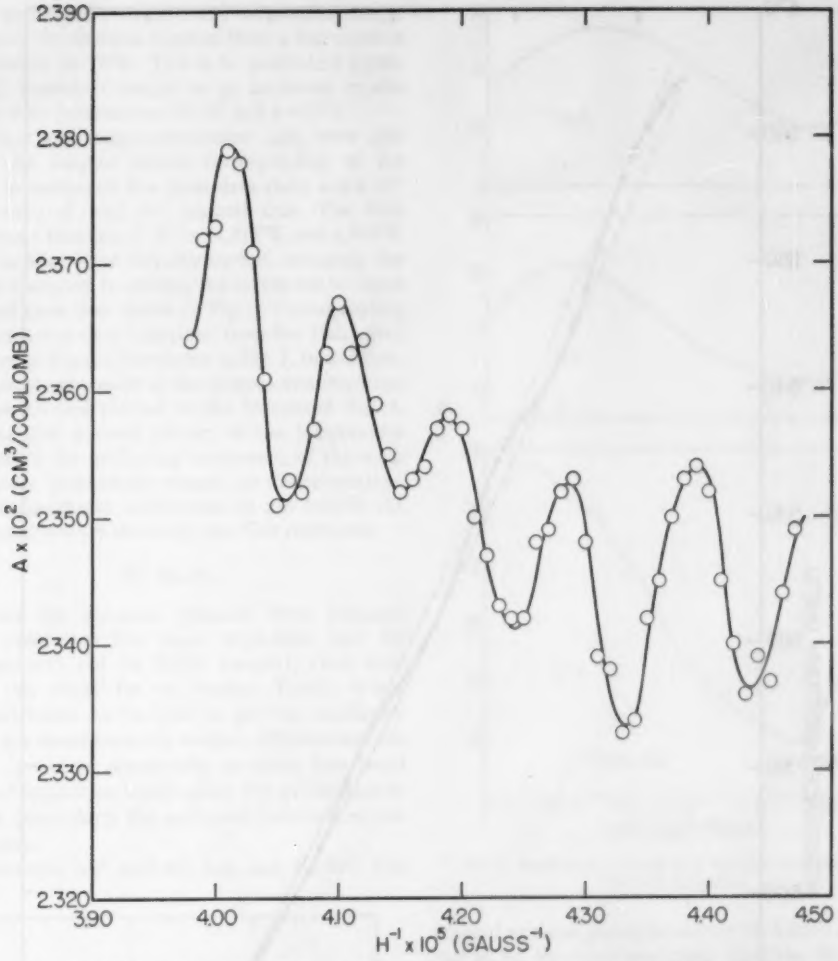


FIG. 2. Field dependence of the Hall coefficient of Sb IV at 1.57°K. H was parallel to the trigonal axis.

total resistance. The dashed lines are the envelope curves drawn through the extremities of the data. At 25 kilogauss the amplitude of the oscillatory components of resistance is 0.3 percent of the total resistance.

Hall coefficient data for Sb IV were also taken at 78°K and 300°K. These results (together with those at 4.21°K) showing A as a function of H are given in Fig. 4. All the data in Fig. 4 were taken with H parallel to the trigonal axis. For this orientation, A is positive (abnormal) over the entire range of temperature and magnetic field strength investigated. The high-field portion of the 4.21°K curve is dashed to indicate that A underwent oscillations in that region. Whereas the 300°K data indicated practically no field dependence, the 78°K data show a definite change in A as a function of H. The sign and magnitude of A at 300°K is in good agreement with the work of Leverton and Dekker.[12]

Data on the variation of resistance as a function of the angle between the trigonal axis and H were taken at 300°K, 78°K, and 4.21°K. The results for Sb IV at $H=25$ kilogauss are shown in Fig. 5 where $\Delta\rho=\rho_H-\rho_0$ and ρ_0 is the resistance in zero field at the particular temperature. Aside from the fact that the minimum in resistance did not occur when H was parallel to the trigonal axis, it can be seen that at 4.21°K, there is a marked departure from the simpler angular dependence exhibited at the higher temperatures. However, it is noted that even in going from 300°K to 78°K, the position of the maximum in resistance shifted by 15° of angle. On the other hand, the position of the minimum

[12] W. F. Leverton and A. J. Dekker, Phys. Rev. **80**, 732 (1950).

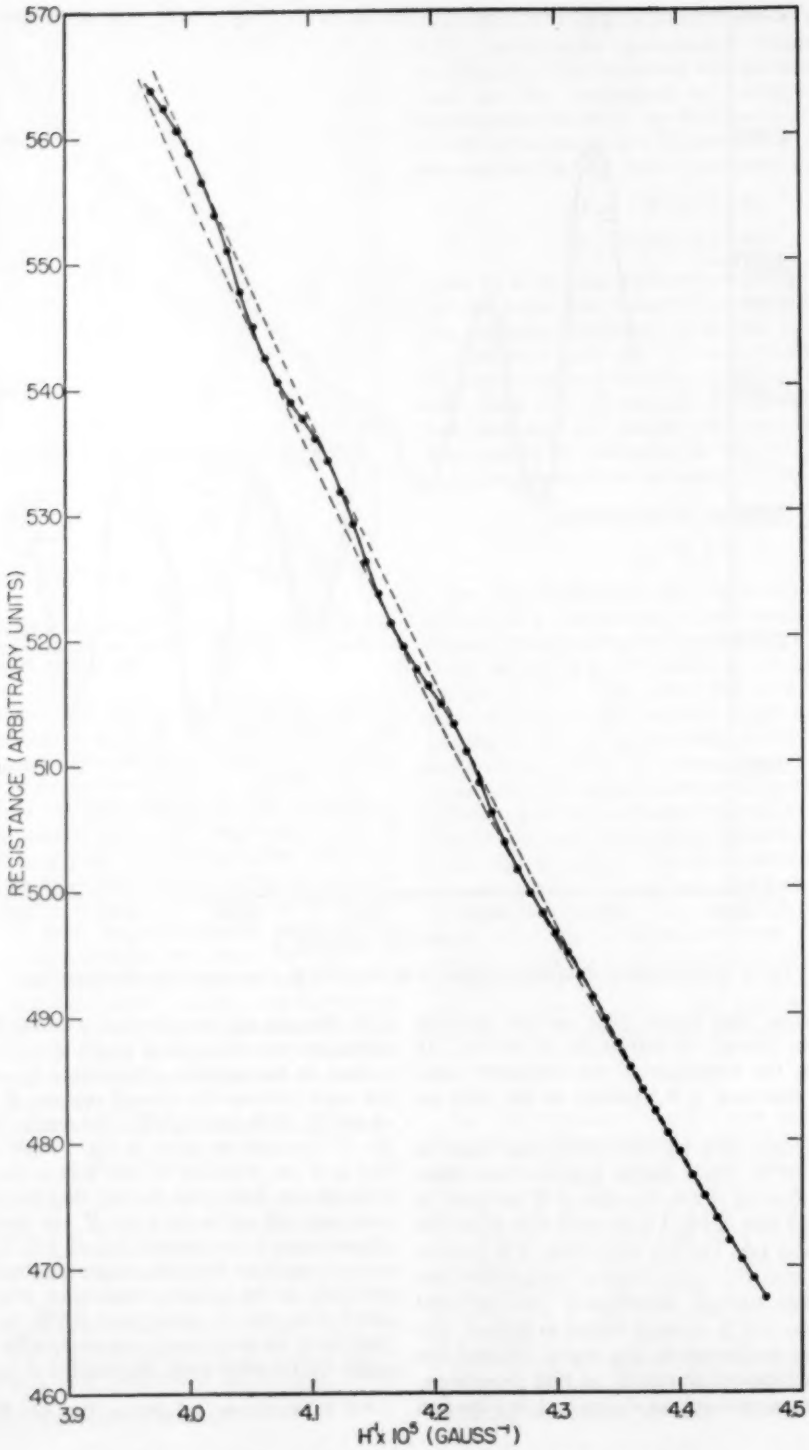


FIG. 3. Magnetoresistance of Sb IV at 1.57°K. H was parallel to the trigonal axis.

remained unchanged over the entire temperature range. Distortion of the rotation diagram from a sine curve is already evident at 78°K. This is in qualitative agreement with Rausch's[11] results on an antimony crystal having $\psi=2.8°$ (whereas our Sb IV had $\psi=15°$).

Hall effect and magnetoresistance data were also taken at the magnet setting corresponding to the minimum in resistance. For these data there was a 15° angle between H and the trigonal axis. The Hall coefficient as a function of H^{-1} at 4.214°K and 1.549°K is shown in Fig. 6 for this orientation. Although the oscillatory character is evident, the curves are far more complicated than that shown in Fig. 2. Corresponding magnetoresistance data [obtained from the Hall effect probes through Eq. (2)] are shown in Fig. 7. In addition, the oscillatory component of the magnetoresistance has been separated and plotted at the bottom of Fig. 7. These data give a vivid picture of the temperature dependence of the oscillating component of the magnetoresistance. Surprisingly enough, for this orientation, the magnetoresistance oscillations do not exhibit the complicated structure shown by the Hall coefficient.

(C) Sb V

This was the specimen prepared from Johnson-Matthey antimony. The major impurities were Pb (0.0275 percent) and As (0.025 percent). Data were taken on this crystal for two reasons. Firstly, it was thought advisable to be able to get the oscillatory effects on specimens from two entirely different sources. Secondly, it was an opportunity to study how small amounts of impurities would affect the galvanometric properties, particularly the oscillatory behavior, at low temperatures.

This specimen had $t=0.685$ mm and $\psi=90°$. The trigonal axis was perpendicular to the face of the plate, just as for the other specimens. Rotation diagrams for the magnetoresistance (at $H=25$ kilogauss) are shown in Fig. 8. It is interesting to note that at all three temperatures the minimum in resistance corresponded to H being parallel to the trigonal axis. Thus, for this particular type of crystal geometry, the position of the trigonal axis could be located from magnetoresistance measurements. Rausch's[11] results at 78°K on an antimony crystal with $\psi=87.8°$ are in good agreement with the present data (even though his data were taken with $H=10.1$ kilogauss).

Hall-effect data were taken with H parallel to the trigonal axis. Figure 9 gives the results at $T=1.549$°K. Although the oscillatory nature of A is evident from this curve, there are several irregularities as compared to the results for Sb IV (shown in Fig. 2). Most striking is the steepness of the envelope curve (dashed) over the range of field shown (22–25 kilogauss). The mean value of A changed by 12 percent while for Sb IV there

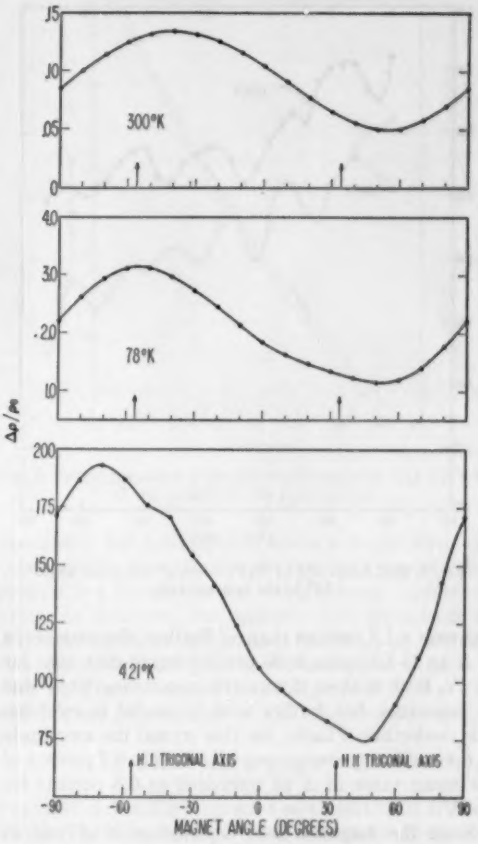


FIG. 5. Resistance of Sb IV as a function of magnet angle ($H=25$ kilogauss).

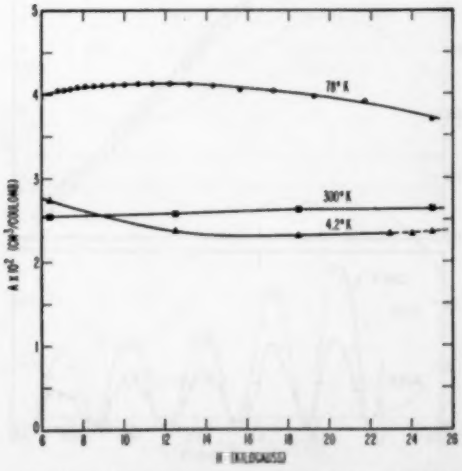


FIG. 4. Field dependence of the Hall coefficient of Sb IV. H was parallel to the trigonal axis.

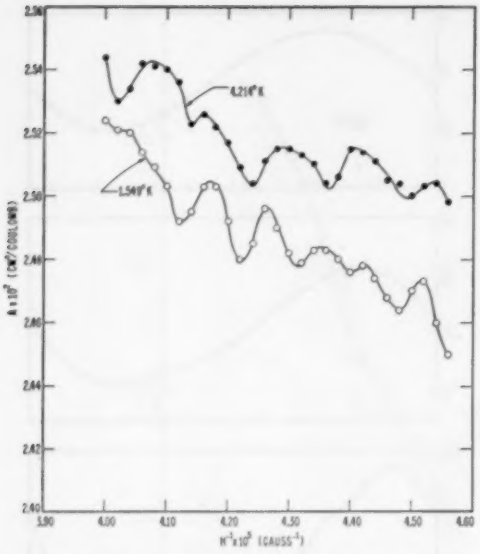


FIG. 6. Hall coefficient of Sb IV when H was at an angle of 15° to the trigonal axis.

was only a 1.5 percent change. Further, the magnitude of A at 25 kilogauss is 60 percent larger than that for Sb IV. Both of these characteristics are most likely due to impurities, but further work is needed to establish this conjecture. Finally, for this crystal the amplitude of the oscillatory component of A is only 0.2 percent of the mean value of A as compared to 0.6 percent for Sb IV.

Since the magnetic field dependence of A was so much stronger for Sb V in high fields, it was thought profitable to make a more detailed study of this behavior over a wider range of field and temperature. The results of such experiments (for Sb V) are given in Fig. 10 where A is plotted as a function of H at three temperatures. H was parallel to the trigonal axis for these data. At 300°K, A is field-independent; at 78°K there appears a field dependence qualitatively similar to that shown in Fig. 4, but the magnitude of A is different for the two cases; at 4.2°K the difference between the two curves is more evident.

DISCUSSION

(A) Period of Oscillations

Sb I magnetoresistance.—From Fig. 1 the principle period of oscillation of the δ curve is 14.1×10^{-7} gauss^{-1}. We shall designate the period of oscillation in H^{-1} as β/E_0 because this is the usual nomenclature in de Haas-van Alphen work. There β is the double effective Bohr magneton (involving effective masses for the electrons) and E_0 is the chemical potential (Fermi energy) of the pertinent de Haas-van Alphen electrons (or holes) measured from the bottom of the significant zone for electrons (or from the top of the zone for holes). From the ellipsoid scheme proposed by Shoenberg[9] for antimony, we find that for our orientation (H perpendicular to the trigonal axis and at an angle of 25° to a binary axis) there would be three values of β/E_0. They are

$$\beta_1/E_0=14.5\times10^{-7}\ \text{gauss}^{-1},$$
$$\beta_2/E_0=\ 7.7\times10^{-7}\ \text{gauss}^{-1},$$
$$\beta_3/E_0=\ 5.9\times10^{-7}\ \text{gauss}^{-1}.$$

The value for β_1/E_0 is in good agreement with our value of 14.1×10^{-7} gauss^{-1}. Our results also indicate that there are subordinate oscillatory terms present, but the data do not warrant making a quantitative comparison to the two other periods of oscillation. Such a detailed study would require another investigation.

Sb IV Hall coefficient and magnetoresistance.—From Fig. 2, β/E_0 for the Hall coefficient oscillations is $9.8\pm0.2\times10^{-7}$ gauss^{-1}. From Fig. 3 β/E_0 for the magnetoresistance oscillations is also $9.8\pm0.2\times10^{-7}$ gauss^{-1}. Thus, within experimental error both these oscillations have the same period. For H parallel to the trigonal axis (the orientation for these measurements) Shoenberg's[9] three oscillatory terms, all have the same period, $\beta/E_0=10.0\times10^{-7}$ gauss^{-1}. Again this is in good agreement with our results. The curve in Fig. 2 suggests

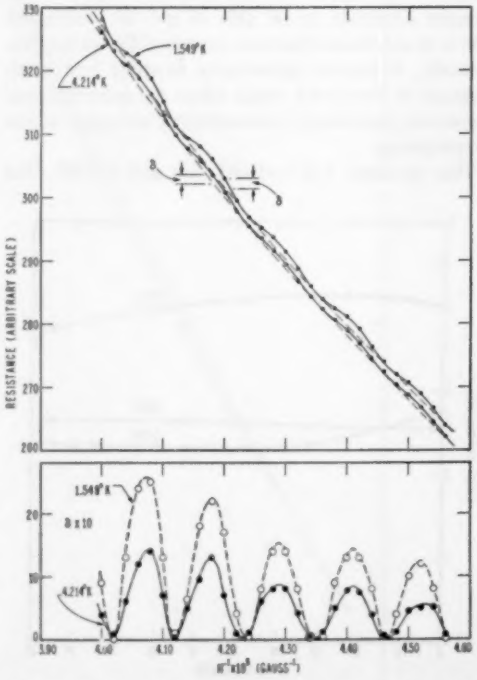


FIG. 7. Magnetoresistance of Sb IV when H was at an angle of 15° to the trigonal axis.

that there may be another oscillatory term present, but this might be due to uncertainties of several degrees in orienting the crystal with respect to H. Peculiarly enough, the magnetoresistance data in Fig. 3 do not show such a modulation of the oscillation amplitude.

(B) Relative Phase of Oscillations

From Figs. 2 and 3, we find that Hall coefficient oscillation maxima occur before the resistance maxima by about 2×10^{-7} gauss^{-1}. There is a good deal of uncertainty about this phase difference because of the arbitrary way in which the envelope curves are drawn and the aspect of beats in Fig. 2. However, the data are given so as to allow some comparison with the results on other materials. In terms of H^{-1}, the magnetoresistance maxima lag the Hall coefficient maxima by roughly 0.4π. Berlincourt and Steele[7] had found that for graphite, magnetoresistance minima lag Hall coefficient minima by about $\pi/4$. Unpublished data of Overton and Berlincourt on bismuth have shown zero phase difference between magnetoresistance and Hall coefficient minima. In light of the different results for three materials no empirical generalization can be made. A detailed model that takes into account the contributions of both holes and electrons is needed to describe each metal individually.

Since the mean value of the Hall coefficient is itself field dependent, it is of interest to see how this is related to the magnetoresistance. A significant function to

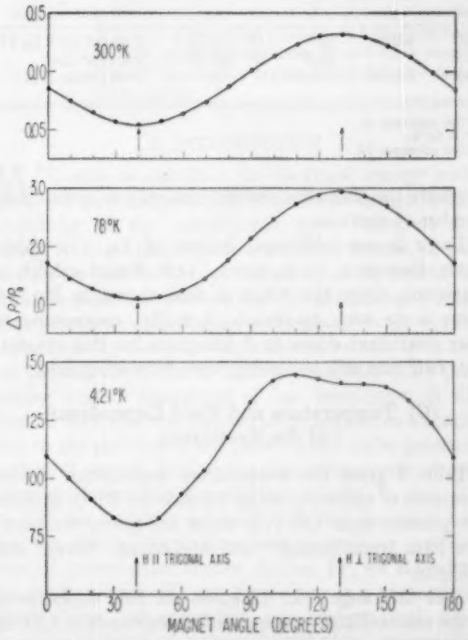


FIG. 8. Resistance of Sb V as a function of magnet angle ($H=25$ kilogauss).

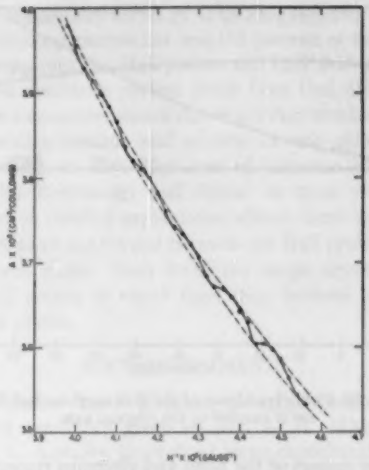


FIG. 9. Field dependence of the Hall coefficient of Sb V at 1.549°K. H was parallel to the trigonal axis.

examine is the product $A\sigma$, where σ is the electrical conductivity. For the simple two-band model, this product is a function of the mobilities and numbers of holes and electrons. The magnetic field dependence of $A\sigma$ might therefore supply more fundamental information than would analyses of the individual variations with H. The voltages, V_A and V_ρ, given in Eqs. (2) and (3), are proportional to AH and $1/\sigma$ respectively, so that

$$V_A/V_\rho \propto A\sigma H. \tag{4}$$

Figure 11 shows V_A/V_ρ plotted as a function of H^{-1} for Sb IV. The data for this graph were taken from the same experiment that gave Figs. 2 and 3. Hence, H was parallel to the trigonal axis and T was 1.57°K. The oscillations in $A\sigma H$ as a function of H^{-1} are clearly separable from the total value. These oscillations are periodic in H^{-1} and have $\beta/E_0=9.5\pm0.2\times10^{-7}$ gauss^{-1}. Another interesting feature of this graph is that the mean value of $A\sigma H$ is quite accurately a linear function of H^{-1} over the range shown in Fig. 11. On the basis of Borovik's[18] analysis of Hall effect and magnetoresistance effect data, one could infer from this linearity that the number of holes are equal to the number of electrons. To show this we follow Borovik and write

$$A\sigma H=\left(\frac{n_1\phi_1^2}{1+\phi_1^2}-\frac{n_2\phi_2^2}{1+\phi_2^2}\right)\Big/\left(\frac{n_1\phi_1}{1+\phi_1^2}+\frac{n_2\phi_2}{1+\phi_2^2}\right), \tag{5}$$

$$\phi_1=eH\tau_1/m_1c; \quad \phi_2=eH\tau_2/m_2c,$$

where $-e$ is the electronic charge; n_1 is the number of holes; n_2 is the number of electrons; m_1 and m_2 are the

[18] E. S. Borovik, J. Exptl. Theoret. Phys. (U.S.S.R.) 23, 83 (1953).

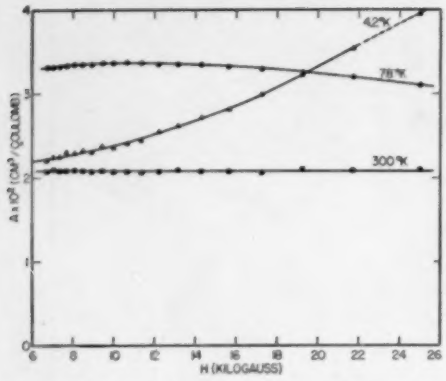


FIG. 10. Field dependence of the Hall coefficient of Sb V for H parallel to the trigonal axis.

effective masses of the holes and electrons respectively, and τ_1 and τ_2 are the respective mean collision times with the lattice. Equation (5) is based on the two-band model of an isotropic metal, whereas antimony is certainly anisotropic. However, it does provide a basis for a qualitative comparison between theory and experiment. The anisotropy would introduce the mass tensor, but the principle functional relations should retain their form. For $\phi_1, \phi_2 \gg 1$ (strong fields) and $n_1 \neq n_2$, Eq. (5) becomes

$$A\sigma H \sim (n_1 - n_2)eH \Big/ c\left(\frac{n_1 m_1}{\tau_1} + \frac{n_2 m_2}{\tau_2}\right). \quad (6)$$

This says that $A\sigma H$ would be increasing linearly with H for strong fields. Our data for Sb (in strong fields) do not show such behavior so we will examine the case $n_1 = n_2$. Then, at any field value, Eq. (5) becomes

$$A\sigma H = (\phi_1 - \phi_2)/(1 + \phi_1\phi_2). \quad (7)$$

For strong fields this can be approximated as

$$A\sigma H \sim c\left(\frac{m_2}{\tau_2} - \frac{m_1}{\tau_1}\right) \Big/ eH. \quad (8)$$

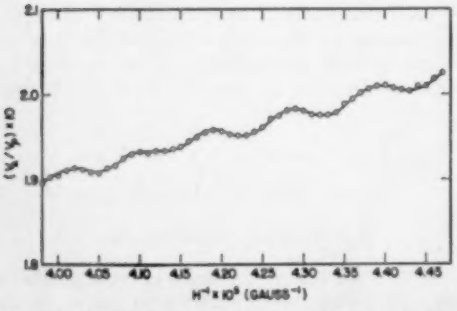


FIG. 11. The field dependence of the ratio of the Hall voltage to the resistance voltage for Sb IV at 1.57°K. H was parallel to the trigonal axis.

Here $A\sigma H$ is a linear function of H^{-1} provided m_2/τ_2 and m_1/τ_1 are field-independent. Since the mean value of $A\sigma H$ in Fig. 11 is a linear function of H^{-1}, we conclude, on the basis of this model, that $n_1 = n_2$ for antimony. However, the oscillatory behavior of $A\sigma H$ necessitates some adjustment to the above development. The simplest assumption one can make consistent with Eq. (8), is that the mobilities of the holes and electrons are oscillatory functions of the magnetic field. On a simple phenomenological development to explain the oscillations of A and σ one can make either the number of carriers and/or their mobilities periodic functions of H^{-1}. However, in the present analysis of the oscillations in $A\sigma H$, a changing number of carriers is of no significance since $n_1 = n_2$ and the numbers do not enter into the resulting expression, Eq. (8). Our conclusion is therefore, that at least the mobilities oscillate with H. Nothing can be said about the variation of n_1 and n_2 with H. A re-examination of the graphite[7] data in light of this analysis has shown that there too $A\sigma H$ undergoes oscillations which are periodic in H^{-1} and that the mean value is decreasing linearly with H^{-1}. Hence, $n_1 = n_2$ would also apply to the case of graphite. The model suggested here differs from that of Levinger and Grimsal[14] in that they attribute the Hall effect oscillations only to a changing number of electrons.

TABLE I. Temperature dependence of electrical resistance of antimony single crystals for $H = 0$. $r_t = \rho_t/\rho_{78°K}$. ϕ = angle between trigonal axis and electric current. ψ = angle between a binary axis and electric current.

$r_{78°K}$	$r_{4.2°K}$	ϕ	ψ	Crystal reference
0.187	0.014	90°	15° (or 75°)	Present data Sb IV
0.188	...	86.3°	88° (or 28°)	Rausch[a] S_{9}[11]
0.260[b]	0.0791	?	?	McLennan *et al.*[c]

[a] See reference 11.
[b] At 82°K.
[c] See reference 15.

There is one additional feature of Eq. (7) which needs discussion. At $\phi_1 \phi_2 = 1$, $A\sigma H$ should exhibit a maximum. Over the range of field shown in Fig. 11 there is no such maximum. Actually, measurements were continued down to 6 kilogauss for this crystal, but $A\sigma H$ was still increasing with decreasing field.

(C) Temperature and Field Dependence of the Resistance

Table I gives the temperature dependence of the resistance of antimony single crystals for $H = 0$. Besides the present data (Sb IV), there are given comparative data from Rausch[11] and McLennan, Niven, and Wilhelm.[15]

Over the range 6–25 kilogauss the field dependence of the nonoscillatory component of resistance at 4.21°K

[14] J. S. Levinger and E. G. Grimsal, Phys. Rev. **94**, 772 (1954).
[15] McLennan, Niven, and Wilhelm, Phil. Mag. **6**, 666 (1928).

is given by the relation

$$\Delta\rho/\rho_0 \propto H^{1.7}, \quad (9)$$

for H either parallel or perpendicular to the trigonal axis. It is noted here that the magnetoresistance of bismuth single crystals[1] has a field dependence very similar to that given by Eq. (9).

Nothing definitive can be said about the temperature or field dependence of the oscillatory component of the magnetoresistance. Figure 7 shows that there is a strong temperature dependence for at least that particular orientation, but for other directions there seems to be little change over the range 1.5°–4.2°K. A more detailed study is needed before any firm conclusions are drawn. There is one more point, however, that is worthy of note. It concerns the fact that the amplitude of the oscillations in magnetoresistance as obtained from the Hall probes is considerably greater (relative to the total resistance) than that obtained from the resistance probes. Specifically for Sb IV at 25 kilogauss and 1.57°K the oscillatory component was 0.3 percent of the total resistance from the Hall probes and only 0.15 percent from the resistance probes. Aside from that difference the two magnetoresistance curves are very similar. They both exhibit maxima and minima at very nearly the same values of H^{-1}. This sort of behavior was also noted by Berlincourt and Steele[7] in their work on graphite. A possible explanation offered there was that the portion of the crystal between the Hall probes may have been purer (both from the single crystal and chemical points of view) than that between the resistance probes.

ACKNOWLEDGMENTS

The author takes pleasure in acknowledging the generous assistance given by Mr. A. H. Mister and Mr. R. C. Anonsen throughout these experiments.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Energy Levels of a Crystal Modified by Alloying or by Pressure

R. H. PARMENTER
RCA Laboratories, Radio Corporation of America, Princeton, New Jersey

(Received May 31, 1955)

A perturbation method of calculating the energy levels of a crystal modified by alloying or by pressure is studied. Initially the calculation is developed by means of conventional first-order perturbation theory. It is then shown what is wrong with this approach and how the calculation can be improved by a modified form of perturbation theory. It is further shown that this modified form of perturbation theory is particularly appropriate to cases where the unperturbed wave functions are expanded in terms of orthogonalized plane waves.

I. INTRODUCTION

WE wish to calculate the electronic energy levels of a crystal modified in a certain fashion from a knowledge of the one-electron wave functions and energy levels of the original unmodified crystal. In particular, we consider the modifications of the crystal (a) under hydrostatic pressure and (b) containing a small mole-percentage of a foreign atom placed in the crystal at random. In the following sections we will work out the general theory for a binary disordered alloy containing a small percentage of one constituent. It will then be seen that the results can be applied as a special case to the problem of the pure crystal under pressure. In Sec. II we will develop the theory by means of conventional first-order perturbation theory. In Sec. III, we will discuss what is wrong with this approach and how our theory can be improved by using a modified form of perturbation theory. In Sec. IV, we show how the results of Sec. III can be written in a particularly simple form by expanding our unperturbed wave functions in terms of orthogonalized plane waves.[1] The special case of a crystal of the diamond structure is worked out in detail. This case is particularly appropriate since the three most important examples of this crystal structure, namely diamond,[2] silicon,[3] and germanium,[4] have all been studied by the orthogonalized plane wave method.

Insofar as it deals with disordered alloys, this paper is intended to supplement a general discussion of the problem in a previous publication by the author.[5] Unlike the previous publication, however, here we will study the formal mathematical aspects of a method suitable for numerical calculations of specific physical systems. In fact, the present study was undertaken with the purpose in mind of forming a basis for a program of detailed numerical calculations on the germanium-silicon alloy and on germanium under hydrostatic

[1] C. Herring, Phys. Rev. **57**, 1169 (1940).

[2] F. Herman, Phys. Rev. **88**, 1210 (1952); **93**, 1214 (1954).

[3] T. O. Woodruff, dissertation, California Institute of Technology, 1955 (unpublished); Phys. Rev. **98**, 1741 (1955).

[4] F. Herman and J. Callaway, Phys. Rev. **89**, 518 (1953); F. Herman, Physica **20**, 801 (1954); Phys. Rev. (to be published).

[5] R. H. Parmenter, Phys. Rev. **97**, 587 (1955).

pressure. Both the alloy system[6] and the pressure system[7] have been studied experimentally recently, and both systems have been discussed from a theoretical standpoint.[8]

It should be emphasized that the methods developed in this paper for handling a disordered alloy are capable of determining how the energy levels of the original host crystal are modified by the introduction of an impurity. These methods cannot, however, determine the *new* energy levels associated with localized states that may occur in the vicinity of each impurity atom.[8] Such localized states will occur when the original host crystal is a semiconductor or insulator and the valence of the impurity atom differs from that of the host atom.

II. CONVENTIONAL PERTURBATION THEORY

Consider the case of a host crystal composed of atoms of type A containing at random lattice sites a small percentage of atoms of type B. Let f be the fraction of lattice sites occupied by B atoms, where $f \ll 1$. In the absence of B atoms, we may express our one-electron crystal potential in the form

$$\sum_n V_a(\mathbf{r}-\mathbf{R}_n), \tag{1}$$

where $V_a(\mathbf{r})$ is an atomic-like potential characteristic of A atoms and the vectors $\mathbf{R}_n$ denote the positions of the atoms of the crystal. In the case where B atoms are present, we follow Nordheim[9] in assuming that the potential may be written as

$$\sum_n V_{a_n}(\mathbf{r}-\alpha\mathbf{R}_n), \tag{2}$$

where $a_n=a$ when $\alpha\mathbf{R}_n$ denotes an atomic site occupied by an A atom, while $a_n=b$ when $\alpha\mathbf{R}_n$ denotes an atomic site occupied by a B atom. α is the ratio of the new lattice constant to the old, and $V_b(\mathbf{r})$ is an atomic-like potential characteristic of B atoms. Since f is small, we may assume that $(\alpha-1)$ is proportional to f (Vegard's law); i.e., we define ϵ and κ such that

$$\epsilon \equiv \alpha - 1 \equiv \kappa f. \tag{3}$$

Since $f \ll 1$, then also $|\epsilon| \ll 1$. At this point we invoke Nordheim's virtual-crystal approximation[9]; i.e., we replace the alloy potential by that portion of it which gives rise to *coherent* electron scattering, namely,

$$\sum_n V(\mathbf{r}-\alpha\mathbf{R}_n), \tag{4}$$

where

$$V(\mathbf{r})=[1-f]V_a(\mathbf{r})+fV_b(\mathbf{r}). \tag{5}$$

As was shown by Muto,[10] the virtual-crystal approximation gives energy levels correctly to the order of conventional first-order perturbation theory.[11] (It is to be remembered that we are dealing with a completely disordered alloy, so that each type of atom is placed on the atom sites in a random fashion.) By the above approximation, we have reduced our problem to one having the periodicity of the lattice.

We wish to solve the Schrödinger equation

$$[-\nabla^2+\textstyle\sum_n V(\mathbf{r}-\alpha\mathbf{R}_n)-E_q'(\mathbf{k})]\Psi_q'(\mathbf{k},\mathbf{r})=0. \tag{6}$$

Here q denotes the particular energy band under consideration, and $\mathbf{k}$ determines the translational properties of the wave function. We are using atomic units. Let $\mathbf{R}_n'$ denote the lattice positions of the Bravais lattice for the undilated lattice; thus $\alpha\mathbf{R}_n'$ will be the corresponding quantities for the dilated lattice. Define

$$\mathfrak{S}_q(\mathbf{k},\mathbf{r})\equiv\Psi_q'(\alpha^{-1}\mathbf{k},\alpha\mathbf{r}). \tag{7}$$

Since

$$\Psi_q'(\mathbf{k},\,\mathbf{r}+\alpha\mathbf{R}_n')=e^{i\mathbf{k}\cdot\alpha\mathbf{R}_n'}\Psi_q'(\mathbf{k},\mathbf{r}), \tag{8}$$

it follows that

$$\mathfrak{S}_q(\mathbf{k},\,\mathbf{r}+\mathbf{R}_n')=e^{i\mathbf{k}\cdot\mathbf{R}_n'}\mathfrak{S}_q(\mathbf{k},\mathbf{r}), \tag{9}$$

so that $\mathfrak{S}_q$ has the correct periodicity properties for the undilated lattice containing only A atoms. Equation (6) may be rewritten

$$[-\alpha^{-2}\nabla^2+\textstyle\sum_n V(\alpha[\mathbf{r}-\mathbf{R}_n])-E_q'(\alpha^{-1}\mathbf{k})]\times\Psi_q'(\alpha^{-1}\mathbf{k},\alpha\mathbf{r})=0,$$

which, after multiplying by α^2, is equivalent to

$$[-\nabla^2+\textstyle\sum_n \alpha^2V(\alpha[\mathbf{r}-\mathbf{R}_n])-\alpha^2E_q'(\alpha^{-1}\mathbf{k})]\times\mathfrak{S}_q(\mathbf{k},\mathbf{r})=0. \tag{10}$$

Let us define

$$V_a'(\mathbf{r})=[2+r(\partial/\partial r)]V_a(\mathbf{r}), \tag{11}$$

$$V_{ba}(\mathbf{r})=V_b(\mathbf{r})-V_a(\mathbf{r}), \tag{12}$$

$$V'(\mathbf{r})=V_a'(\mathbf{r})+\kappa^{-1}V_{ba}(\mathbf{r}). \tag{13}$$

Expand $[\alpha^2V(\alpha\mathbf{r})-V_a(\mathbf{r})]$ in a power series in ϵ, keeping only the leading term since $|\epsilon| \ll 1$. Thus

$$\alpha^2V(\alpha\mathbf{r})-V_a(\mathbf{r})=\epsilon V'(\mathbf{r}). \tag{14}$$

Substituting (14) into (10), we get

$$[-\nabla^2+\textstyle\sum_n V_a(\mathbf{r}-\mathbf{R}_n)+\epsilon\sum_n V'(\mathbf{r}-\mathbf{R}_n)]\mathfrak{S}_q(\mathbf{k},\mathbf{r})=\alpha^2E_q'(\alpha^{-1}\mathbf{k})\mathfrak{S}_q(\mathbf{k},\mathbf{r}). \tag{15}$$

Let $\Psi_q(\mathbf{k},\mathbf{r})$ be a solution of the Schrödinger equation

[6] E. R. Johnson and S. M. Christian, Phys. Rev. **95**, 560 (1954); Levitas, Wang, and Alexander, Phys. Rev. **95**, 846 (1954).

[7] W. Paul and H. Brooks, Phys. Rev. **94**, 1128 (1954); Fan, Shepherd, and Spitzer, Atlantic City Conference on Photoconductivity, November, 1954 (John Wiley and Sons, Inc., New York, 1955); Warschauer, Paul, and Brooks, Phys. Rev. **98**, 1193(A) (1955), and following paper.

[8] One of the most interesting features of these systems is the fact that, as a function of composition or pressure respectively, there appears to be a point where the minimum of the conduction band shifts its position in k-space discontinuously. An explanation of this effect has been proposed by F. Herman, Phys. Rev. **95**, 847 (1954).

[9] L. Nordheim, Ann. Physik **9**, 607 and 641 (1931).

[10] T. Muto, Sci. Papers Inst. Phys. Chem. Research (Tokyo) **34**, 377 (1938).

[11] For a discussion of the effect on the energy levels of higher orders of perturbation theory, see reference 5.

appropriate to the crystal containing only A atoms,[12] i.e.,

$$[-\nabla^2+\sum_n V_a(\mathbf{r}-\mathbf{R}_n)]\Psi_q(\mathbf{k},\mathbf{r})=E_q(\mathbf{k})\Psi_q(\mathbf{k},\mathbf{r}). \quad (16)$$

We see that we may solve Eq. (15) by means of first-order perturbation theory, taking the Ψ's as our unperturbed wave functions. Since the perturbation potential

$$\sum_n V'(\mathbf{r}-\mathbf{R}_n)$$

in no way affects the symmetry of the unperturbed host crystal, then we are justified in using *nondegenerate* perturbation theory, the perturbation potential being unable to remove any degeneracies associated with symmetry. Thus

$$\alpha^2E_q'(\alpha^{-1}\mathbf{k})=E_q(\mathbf{k})+(\alpha-1)\mathbf{V}_q'(\mathbf{k}), \quad (17)$$

or

$$E_q'(\mathbf{k})=\alpha^{-2}[E_q(\alpha\mathbf{k})+(\alpha-1)\mathbf{V}_q'(\alpha\mathbf{k})], \quad (18)$$

where

$$\mathbf{V}_q'(\mathbf{k})=\frac{\int_\infty \Psi_q^*(\mathbf{k},\mathbf{r})\sum_n V'(\mathbf{r}-\mathbf{R}_n)\Psi_q(\mathbf{k},\mathbf{r})d\tau}{\int_\infty \Psi_q^*(\mathbf{k},\mathbf{r})\Psi_q(\mathbf{k},\mathbf{r})d\tau}. \quad (19)$$

($\int_\infty d\tau$ denotes an integration over the entire volume of the crystal.) Expanding Eq. (18) in a power series in ϵ, keeping only the constant and linear terms, we get

$$E_q'(\mathbf{k})=E_q(\mathbf{k})+\epsilon[2\mathbf{k}\cdot\mathbf{v}_q(\mathbf{k})-2E_q(\mathbf{k})+\mathbf{V}_q'(\mathbf{k})]f, \quad (20)$$

where the velocity $\mathbf{v}_q(\mathbf{k})$ is given by

$$\mathbf{v}_q(\mathbf{k})=\tfrac{1}{2}\nabla_k E_q(\mathbf{k}). \quad (21)$$

In Eq. (13), the definition of $V'(\mathbf{r})$, the term $\kappa^{-1}V_{bn}$ arises from the presence of type B atoms in the lattice. The term V_a', however, arises solely from the change in lattice spacing. Similarly, in Eq. (18) the term $\alpha^{-2}E_q(\alpha\mathbf{k})$ arises solely from the change in lattice spacing. Thus by specializing to the case where $\kappa^{-1}=0$ and $f=0$ such that ϵ remains finite (i.e., by removing the term $\kappa^{-1}V_{bn}$), we obtain the effect of dilation or compression on the energy levels of a perfect crystal. An amusing example of the effect (or lack of effect) of dilation is the case of the empty lattice where all potentials vanish. The solution to the Schrödinger equation is a plane wave, while

$$E_q(\mathbf{k})=k^2.$$

Clearly the energy levels must be independent of any hypothetical dilation, on physical grounds. To see this mathematically, we note that

$$E_q'(\mathbf{k})=\alpha^{-2}E_q(\alpha\mathbf{k})=k^2=E_q(\mathbf{k}).$$

III. MODIFIED PERTURBATION THEORY

In essence, the procedure carried out in the previous section is equivalent to that of *scaling* the wave function appropriate to the undilated host crystal, this scaling to be done in both r- and k-space, and then taking the expectation value of the Hamiltonian appropriate to the alloy with respect to this scaled wave function. There are two objections to this procedure. The first objection is that we have effectively assumed that the square of the wave function appropriate to the alloy has the periodicity of the Bravais lattice, an assumption which is certainly not true.[13] It would be desirable to construct some approximate wave function appropriate to the disordered alloy which correctly expresses the fact that the electronic charge density in the neighborhood of a B atom is different from that in the neighborhood of an A atom. The second objection has to do with the fact that when the lattice constant of a crystal is changed, the change in the wave function cannot be represented by scaling alone. In fact, in the immediate vicinity of a particular ion core of the crystal, the wave function, rather than being scaled, probably remains unchanged (aside from a phase factor) since as a first approximation the crystal potential remains unchanged (aside from an additive constant). If we had assumed that the crystal potential is scaled upon changing the lattice constant, then Eq. (2) would have been replaced by

$$\sum_n V_{an}(\alpha^{-1}[\mathbf{r}-\alpha\mathbf{R}_n]). \quad (22)$$

It is felt by the writer, however, that Eq. (2) is a better approximation to the crystal potential for the alloy than in Eq. (22).

Both of the above objections can be overcome by modifying the perturbation procedure in the following fashion. First we consider the undilated host crystal containing only A atoms. We designate by $\Phi_s(\mathbf{k},\mathbf{r})$ an ion-core Bloch wave function appropriate to this case, and by $\Psi_q(\mathbf{k},\mathbf{r})$ any of the other Bloch wave functions appropriate to this case. Since the Ψ's and Φ's are solutions to the same Schrödinger equation, the Ψ's must be orthogonal to the Φ's. In analogy with Herring's method of defining an orthogonalized plane wave,[1] we consider a function $\Gamma_q(\mathbf{k},\mathbf{r})$ such that

$$\Psi_q(\mathbf{k},\mathbf{r})=\Gamma_q(\mathbf{k},\mathbf{r})-\sum_s \Phi_s(\mathbf{k},\mathbf{r})\times\int \Phi_s^*(\mathbf{k},\mathbf{r})\Gamma_q(\mathbf{k},\mathbf{r})d\tau. \quad (23)$$

[12] With regard to notation, throughout this paper we will prime wave functions appropriate to the disordered alloy and leave unprimed wave functions appropriate to the unperturbed crystal containing only A atoms. Also we will use Greek and German symbols to denote exact and approximate wave functions respectively, appropriate to either the unperturbed or the perturbed crystal; e.g., $\mathfrak{S}$ is an approximate solution of Eq. (16) appropriate to the unperturbed cystal.

[13] It is this incorrect assumption which leads to the complete inability of conventional perturbation theory to correctly predict the ion-core energy levels of a disordered alloy, a difficulty discussed in reference 5.

Assuming the orthonormality of the Φ's,

$$\int_\infty \Phi_s^*(\mathbf{k},\mathbf{r})\Phi_t(\mathbf{k},\mathbf{r})d\tau=\delta_{s,t}, \tag{24}$$

we see that Eq. (23) merely expresses the fact that the Ψ's are orthogonal to the Φ's, i.e.,

$$\int_\infty \Phi_s^*(\mathbf{k},\mathbf{r})\Psi_q(\mathbf{k},\mathbf{r})d\tau=0. \tag{25}$$

In a case of a monatomic host crystal where there is more than one atom per unit cell of the crystal, the index s appearing in the last three equations really stands for two indices, s and σ, where s denotes the type of atomic orbital from which the Bloch function is composed, while σ denotes with which atom in the unit cell the Bloch function is associated; i.e.,

$$\Phi_{s,\sigma}(\mathbf{k},\mathbf{r})=N^{-\frac{1}{2}}\sum_n \exp[i\mathbf{k}\cdot(\mathbf{R}_n'+\mathfrak{R}_\sigma)] \times\varphi_s(\mathbf{r}-\mathbf{R}_n'-\mathfrak{R}_\sigma). \tag{26}$$

The vector $\mathfrak{R}_\sigma$ denotes the position of the σth atom of the unit cell with respect to the center of the unit cell. As in the previous section, the Bravais lattice vectors $\mathbf{R}_n'$ denote the positions of the centers of the unit cells, while the atomic positions, previously denoted by $\mathbf{R}_n$, can be written as $\mathbf{R}_n'+\mathfrak{R}_\sigma$. N denotes the number of unit cells in the complete crystal. Let us take the ion-core atomic orbitals on a given atomic site to be orthonormal and assume that the atomic orbitals on neighboring atomic sites do not overlap; i.e.,

$$\int_\infty \varphi_s^*(\mathbf{r}-\mathbf{R}_m'-\mathfrak{R}_\sigma)\varphi_t(\mathbf{r}-\mathbf{R}_n'-\mathfrak{R}_\tau)d\tau =\delta_{s,t}\delta_{m,n}\delta_{\sigma,\tau}. \tag{27}$$

We see that Eq. (24) is consistent with Eqs. (26) and (27). Using the fact that

$$\Gamma_q(\mathbf{k},\mathbf{r}+\mathbf{R}_n')=e^{i\mathbf{k}\cdot\mathbf{R}_n'}\Gamma_q(\mathbf{k},\mathbf{r}), \tag{28}$$

we see that Eq. (23) can be rewritten

$$\Psi_q(\mathbf{k},\mathbf{r})=\Gamma_q(\mathbf{k},\mathbf{r})-\sum_{s,n}\varphi_s(\mathbf{r}-\mathbf{R}_n) \times\int_\infty \varphi_s^*(\mathbf{r}-\mathbf{R}_n)\Gamma_q(\mathbf{k},\mathbf{r})d\tau. \tag{29}$$

Equation (29) expresses the fact that the Ψ's are orthogonal to the φ's, i.e.,

$$\int_\infty \varphi_s^*(\mathbf{r}-\mathbf{R}_n)\Psi_q(\mathbf{k},\mathbf{r})d\tau=0. \tag{30}$$

This is consistent with the fact that the φ's, rather than the Φ's, may be considered to be the ion-core solutions to the Schrödinger equation for the undilated host crystal. Although such a point of view is not particularly advantageous for a perfect crystal, it will be very advantageous when we consider the disordered alloy. By means of Eq. (29), we have broken up the wave function $\Psi_q(\mathbf{k},\mathbf{r})$ into two parts, one part

$$-\sum_{s,n}\varphi_s(\mathbf{r}-\mathbf{R}_n)\int_\infty \varphi_s^*(\mathbf{r}-\mathbf{R}_n)\Gamma_q(\mathbf{k},\mathbf{r})d\tau \tag{31}$$

being appreciable only in the immediate vicinity of each ion core of the crystal and representing the rapid variations of the wave function in these regions, while the other part $\Gamma_q(\mathbf{k},\mathbf{r})$ is more smoothly varying and is of importance throughout all regions of the crystal.

Turning now to the disordered alloy, we define the function

$$\mathfrak{S}_q'(\mathbf{k},\mathbf{r})\equiv\mathfrak{G}_q'(\mathbf{k},\mathbf{r})-\sum_{s,n}\varphi_{s,a_n}(\mathbf{r}-\alpha\mathbf{R}_n) \times\int_\infty \varphi_{s,a_n}^*(\mathbf{r}-\alpha\mathbf{R}_n)\mathfrak{G}_q'(\mathbf{k},\mathbf{r})d\tau, \tag{32}$$

where

$$\mathfrak{G}_q'(\mathbf{k},\mathbf{r})\equiv\Gamma_q(\alpha\mathbf{k},\alpha^{-1}\mathbf{r}). \tag{33}$$

As in the previous section, the index a_n appearing in Eq. (32) designates the type of atom appearing on the nth atomic site. In analogy with Eq. (28), we have

$$\mathfrak{G}_q'(\mathbf{k},\mathbf{r}+\alpha\mathfrak{R}_n')=e^{i\mathbf{k}\cdot\alpha\mathbf{R}_n'}\mathfrak{G}_q'(\mathbf{k},\mathbf{r}), \tag{34}$$

so that this portion of the wave function $\mathfrak{S}_q'(\mathbf{k},\mathbf{r})$ has been scaled from the corresponding portion of the wave function Ψ_q appropriate to the undilated host crystal. The remaining portion of the wave function $\mathfrak{S}_q'$, namely

$$-\sum_{s,n}\varphi_{s,a_n}(\mathbf{r}-\alpha\mathbf{R}_n)\int_\infty \varphi_{s,a_n}^*(\mathbf{r}-\alpha\mathbf{R}_n)\mathfrak{G}_q'(\mathbf{k},\mathbf{r})d\tau, \tag{35}$$

has not been obtained by scaling (31), however, but by replacing (31) by linear combinations of the ion-core wave functions appropriate to the modified crystal, their coefficients being chosen such that the $\mathfrak{S}_q''$s so obtained are orthogonal to these latter ion-core wave functions,[14] i.e.,

$$\int_\infty \varphi_{s,a_n}^*(\mathbf{r}-\alpha\mathbf{R}_n)\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau=0. \tag{36}$$

Equation (36) is the analogy for the modified crystal to Eq. (30), which holds for the unmodified crystal. We wish to approximate the wave function appropriate to the disordered alloy by $\mathfrak{S}_q'(\mathbf{k},\mathbf{r})$. It is apparent that such a procedure in large measure removes the two basic defects of the conventional perturbation procedure

[14] The writer is indebted to F. Herman for pointing out the importance of working with approximate alloy wave functions which are orthogonal to the ion-core states of the alloy.

which were discussed at the beginning of this section. It should be mentioned that $\mathbf{k}$ is no longer a good quantum number, since unlike Eq. (8)

$$\mathfrak{S}_q'(\mathbf{k}, \mathbf{r}+\alpha\mathbf{R}_n') = e^{i\mathbf{k}\cdot\alpha\mathbf{R}_n'}\mathfrak{S}_q'(\mathbf{k},\mathbf{r}) \tag{37}$$

does not hold in general. It does hold, however, in regions outside of the ion cores of the alloy where (35) vanishes. The method expounded in this section therefore treats $\mathbf{k}$ as a certain approximation to a good quantum number. That such should be the case is, of course, suggested by the results of conventional perturbation theory given in Sec. II.

We shall approximate the energy for the disordered alloy by

$$E_q'(\mathbf{k}) = \frac{\int_\infty \mathfrak{S}_q'^*(\mathbf{k},\mathbf{r})H'\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau}{\int_\infty \mathfrak{S}_q'^*(\mathbf{k},\mathbf{r})\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau}, \tag{38}$$

the expectation value of H' with respect to the approximate wave function $\mathfrak{S}_q'$. As we shall see presently, $E_q'(\mathbf{k})$ is the same for all random configurations of atoms in the alloy consistent with the specified composition. The Hamiltonian H' is given by

$$H' = -\nabla^2 + \textstyle\sum_n V_{a_n}(\mathbf{r}-\alpha\mathbf{R}_n) - \mathbf{V}_0'. \tag{39}$$

The additive constant

$$\mathbf{V}_0' \equiv \textstyle\sum_n{}' V(\alpha\mathbf{R}_n - \alpha\mathbf{R}_m) \tag{40}$$

has been incorporated into the Hamiltonian for convenience in the calculation. $V(\mathbf{r})$ is the potential defined by Eq. (5), and the prime over the summation in Eq. (40) indicates that the term for $\mathbf{R}_n = \mathbf{R}_m$ is omitted. (We shall restrict ourselves to crystals having sufficient symmetry for $\mathbf{V}_0'$ to be independent of the index m.) Aside from the constant $\mathbf{V}_0'$, the crystal potential in H' is that given by Eq. (2). Since in the immediate neighborhood of the nth atomic site of the crystal, the crystal potential may be closely approximate (aside from an additive constant) by the atomic-like potential centered on that site, we shall define the φ's such that

$$[-\nabla^2 + V_a(\mathbf{r})]\varphi_{s,a}(\mathbf{r}) = E_{s,a}\varphi_{s,a}(\mathbf{r}), \tag{41}$$

and similarly for $\varphi_{s,b}$.[15] Thus, to a good approximation,

$$H'\varphi_{s,a_n}(\mathbf{r}-\alpha\mathbf{R}_n) = \{E_{s,a_n} - \mathbf{V}_0' + \textstyle\sum_m{}' V_{a_m}(\alpha[\mathbf{R}_m - \mathbf{R}_n])\} \times \varphi_{s,a_n}(\mathbf{r}-\alpha\mathbf{R}_n) \tag{42}$$

(the prime over the summation indicating that the term for $m=n$ is omitted). We now find

$$\begin{aligned}\int_\infty \mathfrak{S}_p'^*(\mathbf{k},\mathbf{r})H'\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau &= \int_\infty \mathfrak{B}_p'^*(\mathbf{k},\mathbf{r})H_0'\mathfrak{B}_q'(\mathbf{k},\mathbf{r})d\tau \\ &\quad - N\sum_{s,a}(1-f)E_{s,a}\int_\infty \varphi_{s,a}(\mathbf{r}-\alpha\mathfrak{R}_s)\mathfrak{B}_p'^*(\mathbf{k},\mathbf{r})d\tau \\ &\quad \times \int_\infty \varphi_{s,a}^*(\mathbf{r}-\alpha\mathfrak{R}_s)\mathfrak{B}_q'(\mathbf{k},\mathbf{r})d\tau \\ &\quad - N\sum_{s,a} fE_{s,b}\int_\infty \varphi_{s,b}(\mathbf{r}-\alpha\mathfrak{R}_s)\mathfrak{B}_p'^*(\mathbf{k},\mathbf{r})d\tau \\ &\quad \times \int_\infty \varphi_{s,b}^*(\mathbf{r}-\alpha\mathfrak{R}_s)\mathfrak{B}_q'(\mathbf{k},\mathbf{r})d\tau, \end{aligned} \tag{43}$$

where H_0' is defined as

$$H_0' = -\nabla^2 + \textstyle\sum_n V(\mathbf{r}-\alpha\mathbf{R}_n) - \mathbf{V}_0'. \tag{44}$$

Similarly

$$\begin{aligned}\int_\infty \mathfrak{S}_p'^*(\mathbf{k},\mathbf{r})\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau &= \int_\infty \mathfrak{B}_p'^*(\mathbf{k},\mathbf{r})\mathfrak{B}_q'(\mathbf{k},\mathbf{r})d\tau \\ &\quad - N\sum_{s,a}(1-f)\int_\infty \varphi_{s,a}(\mathbf{r}-\alpha\mathfrak{R}_s)\mathfrak{B}_p'^*(\mathbf{k},\mathbf{r})d\tau \\ &\quad \times \int_\infty \varphi_{s,a}^*(\mathbf{r}-\alpha\mathfrak{R}_s)\mathfrak{B}_q'(\mathbf{k},\mathbf{r})d\tau \\ &\quad - N\sum_{s,a} f\int_\infty \varphi_{s,b}(\mathbf{r}-\alpha\mathfrak{R}_s)\mathfrak{B}_p'^*(\mathbf{k},\mathbf{r})d\tau \\ &\quad \times \int_\infty \varphi_{s,b}^*(\mathbf{r}-\alpha\mathfrak{R}_s)\mathfrak{B}_q'(\mathbf{k},\mathbf{r})d\tau. \end{aligned} \tag{45}$$

For the special case where the band indices p and q are the same, Eqs. (43), (44), and (45) prove the statement made previously that $E_q'(\mathbf{k})$ is independent of the configuration of atoms of the alloy. Let $\Psi_p(\alpha\mathbf{k},\alpha^{-1}\mathbf{r})$ and and $\Psi_q(\alpha\mathbf{k},\alpha^{-1}\mathbf{r})$ be two degenerate wave functions associated with the undilated host crystal, with $\alpha\mathbf{k}$ being a symmetry point in k-space, Ψ_p and Ψ_q having different point symmetries in real space. If $\mathfrak{S}_p'(\mathbf{k},\mathbf{r})$ and $\mathfrak{S}_q'(\mathbf{k},\mathbf{r})$ are the corresponding wave functions for the alloy, then by applying symmetry considerations to Eqs. (43), (44), and (45), we find that

$$\int_\infty \mathfrak{S}_p'^*(\mathbf{k},\mathbf{r})H'\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau = 0, \tag{46}$$

and

$$\int \mathfrak{S}_p'^*(\mathbf{k},\mathbf{r})\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau = 0. \tag{47}$$

[15] We shall assume that the atomic-like potentials V_a and V_b may be approximated by the corresponding isolated-atom potentials. Thus the φ_s's and the E_s's are those for the isolated atoms.

It is for this reason that we are justified in using a modified form of *nondegenerate* first-order perturbation theory.

For the Hamiltonian of the undilated host crystal we take

$$H=-\nabla^2+\sum_n V_a(\mathbf{r}-\mathbf{R}_n)-\mathbf{V}_0, \tag{48}$$

where

$$\mathbf{V}_0\equiv\sum_n{}' V_a(\mathbf{R}_n-\mathbf{R}_m). \tag{49}$$

The corresponding energy is

$$E_q(\mathbf{k})=[I_q(\mathbf{k})]^{-1}\int_\infty \Psi_q^*(\mathbf{k},\mathbf{r})H\Psi_q(\mathbf{k},\mathbf{r})d\tau, \tag{50}$$

where

$$I_q(\mathbf{k})\equiv\int_\infty \Psi_q^*(\mathbf{k},\mathbf{r})\Psi_q(\mathbf{k},\mathbf{r})d\tau=\int_\infty|\Gamma_q(\mathbf{k},\mathbf{r})|^2d\tau -N\sum_{s,\sigma}\left|\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2, \tag{51}$$

and

$$\int_\infty \Psi_q^*(\mathbf{k},\mathbf{r})H\Psi_q(\mathbf{k},\mathbf{r})d\tau =\int_\infty\Gamma_q^*(\mathbf{k},\mathbf{r})H\Gamma_q(\mathbf{k},\mathbf{r})d\tau -N\sum_{s,\sigma}E_{s,a}\left|\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2. \tag{52}$$

We now define

$$\varphi_{s,a}'(\mathbf{r})\equiv r\frac{\partial}{\partial r}\varphi_{s,a}(\mathbf{r}). \tag{53}$$

Making use of Eqs. (33), (45), (51), and (53), and keeping only terms linear in $\epsilon=\alpha-1$, we get

$$\begin{aligned}\alpha^{-3}&\int_\infty \mathfrak{S}_q'^*(\mathbf{k},\mathbf{r})\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau\\ &-\int_\infty\Psi_q^*(\alpha\mathbf{k},\mathbf{r})\Psi_q(\alpha\mathbf{k},\mathbf{r})d\tau\\ =&-\epsilon N\sum_{s,\sigma}\Big\{3\left|\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2\\ &+\kappa^{-1}\left|\int_\infty\varphi_{s,b}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2\\ &-\kappa^{-1}\left|\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2\\ &+2\,\mathrm{Re}\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\\ &\times\int_\infty\varphi_{s,a}'(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q^*(\mathbf{k},\mathbf{r})d\tau\Big\}.\end{aligned} \tag{54}$$

where Re means "real part of." In a similar fashion, we have

$$\begin{aligned}\alpha^{-3}&\int_\infty \mathfrak{S}_q'^*(\mathbf{k},\mathbf{r})H'\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau\\ &-\alpha^{-2}\int_\infty\Psi_q^*(\alpha\mathbf{k},\mathbf{r})H\Psi_q(\alpha\mathbf{k},\mathbf{r})d\tau\\ =&\epsilon\int_\infty|\Gamma_q(\mathbf{k},\mathbf{r})|^2[\sum_n V'(\mathbf{r}-\mathbf{R}_n)-\sum_n{}' V'(\mathbf{R}_n-\mathbf{R}_m)]d\tau\\ &-\epsilon N\sum_{s,\sigma}\Big\{5E_{s,a}\left|\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2\\ &+\kappa^{-1}E_{s,b}\left|\int_\infty\varphi_{s,b}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2\\ &-\kappa^{-1}E_{s,a}\left|\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2\\ &+2E_{s,a}\,\mathrm{Re}\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathbf{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\\ &\times\int_\infty\varphi_{s,a}'(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q^*(\mathbf{k},\mathbf{r})d\tau\Big\}.\end{aligned} \tag{55}$$

Substituting Eqs. (50), (51), (54), and (55) into Eq. (38), we get

$$E_q'(\mathbf{k})-\alpha^{-2}E_q(\alpha\mathbf{k})=\epsilon[I_q(\mathbf{k})]^{-1}\{[A_{1q}(\mathbf{k})+A_{2q}(\mathbf{k})] +\kappa^{-1}[B_{1q}(\mathbf{k})+B_{2q}(\mathbf{k})]\}, \tag{56}$$

where

$$A_{1q}(\mathbf{k})\equiv\int_\infty|\Gamma_q(\mathbf{k},\mathbf{r})|^2[\sum_n V_a'(\mathbf{r}-\mathbf{R}_n) -\sum_n{}' V_a'(\mathbf{R}_n-\mathbf{R}_m)]d\tau, \tag{57}$$

$$\begin{aligned}A_{2q}(\mathbf{k})\equiv N\sum_{s,\sigma}\Big\{&[3E_q(\mathbf{k})-5E_{s,a}]\\ &\times\left|\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2\\ &+2[E_q(\mathbf{k})-E_{s,a}]\,\mathrm{Re}\int_\infty\varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\\ &\times\Gamma_q(\mathbf{k},\mathbf{r})d\tau\int_\infty\varphi_{s,a}'(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q^*(\mathbf{k},\mathbf{r})d\tau\Big\},\end{aligned} \tag{58}$$

$$B_{1q}(\mathbf{k})\equiv\int_\infty|\Gamma_q(\mathbf{k},\mathbf{r})|^2[\sum_n V_{ba}(\mathbf{r}-\mathbf{R}_n) -\sum_n{}' V_{ba}(\mathbf{R}_m-\mathbf{R}_m)]d\tau, \tag{59}$$

$$B_{2q}(\mathbf{k})\equiv N\sum_{s,\sigma}\Big\{[E_q(\mathbf{k})-E_{s,b}]\Big|\int_\omega \varphi_{s,b}^*(\mathbf{r}-\mathfrak{R}_\sigma)\times\Gamma_q(\mathbf{k},\mathbf{r})d\tau\Big|^2-[E_q(\mathbf{k})-E_{s,a}]\times\Big|\int_\omega \varphi_{s,a}^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\Big|^2\Big\}. \quad (60)$$

Making use of Eqs. (3) and (21), and defining

$$A_{2q}(\mathbf{k})\equiv 2[\mathbf{k}\cdot\mathbf{v}_q(\mathbf{k})-E_q(\mathbf{k})]I_q(\mathbf{k}), \quad (61)$$

we finally obtain

$$E_q'(\mathbf{k})-E_q(\mathbf{k})=\kappa[I_q(\mathbf{k})]^{-1}[A_{1q}(\mathbf{k})+A_{2q}(\mathbf{k})+A_{3q}(\mathbf{k})]f+[I_q(\mathbf{k})]^{-1}[B_{1q}(\mathbf{k})+B_{2q}(\mathbf{k})]f. \quad (62)$$

It is clear that the effect of a dilation upon a perfect crystal may be studied with the aid of Eq. (62) simply by removing the B terms.

IV. APPLICATION TO THE DIAMOND STRUCTURE

The procedure developed in the preceding section is particularly suited to an unmodified crystal for which wave functions and energy levels have been calculated by the orthogonalized plane wave (OPW) method.[1] In this method the wave function is expanded in terms of orthogonalized pane waves $X(\mathbf{k}_i,\mathbf{r})$, $\mathbf{k}_i$ being defined as

$$\mathbf{k}_i\equiv\mathbf{k}+\mathbf{K}_i, \quad (63)$$

$\mathbf{K}_i$ being a vector of the reciprocal lattice such that

$$e^{i\mathbf{K}_i\cdot\mathbf{R}_n'}=1, \quad (64)$$

for all i and n. $X(\mathbf{k}_i,\mathbf{r})$ is defined as

$$X(\mathbf{k}_i,\mathbf{r})\equiv(N\Omega)^{-\frac{1}{2}}e^{i\mathbf{k}_i\cdot\mathbf{r}}-\Omega^{-1}\sum_{s,\sigma}\exp(i\mathbf{K}_i\cdot\mathfrak{R}_\sigma)\times\mu_{s,a}(\mathbf{k}_i)\Phi_{s,\sigma,a}(\mathbf{k},\mathbf{r}), \quad (65)$$

where

$$\mu_{s,a}(\mathbf{k}_i)=N^{-\frac{1}{2}}\exp(-i\mathbf{K}_i\cdot\mathfrak{R}_\sigma)\times\int_\omega e^{i\mathbf{k}_i\cdot\mathbf{r}}\Phi_{s,\sigma,a}^*(\mathbf{k},\mathbf{r})d\tau=\int_\omega e^{i\mathbf{k}_i\cdot\mathbf{r}}\varphi_{s,a}^*(\mathbf{r})d\tau, \quad (66)$$

and where Ω is the volume of the unit cell of the unmodified crystal. We see now that

$$\int_\omega \Phi_{s,\sigma,a}^*(\mathbf{k},\mathbf{r})X(\mathbf{k}_i,\mathbf{r})d\tau=0. \quad (67)$$

Expanding $\Psi_q(\mathbf{k},\mathbf{r})$ in terms of the X's, we have

$$\Psi_q(\mathbf{k},\mathbf{r})=\sum_i \mathfrak{A}_q(\mathbf{k}_i)X(\mathbf{k}_i,\mathbf{r}). \quad (68)$$

Hence it follows that $\Gamma_q(\mathbf{k},\mathbf{r})$, as defined by Eq. (23), can be written as

$$\Gamma_q(\mathbf{k},\mathbf{r})=(N\Omega)^{-\frac{1}{2}}\sum_i \mathfrak{A}_q(\mathbf{k}_i)e^{i\mathbf{k}_i\cdot\mathbf{r}}. \quad (69)$$

It is because of the simplicity of Eq. (69) that the orthogonalized plane wave method is particularly adapted to the results of Sec. III.

We shall now specialize to the case where the unmodified crystal has the diamond structure, this being a case of great practical importance. There are now two atoms per unit cell such that $\mathfrak{R}_\sigma=\pm\mathfrak{R}$, where $\mathfrak{R}$ points in the (111) direction, and $2\mathfrak{R}$ is the nearest-neighbor distance. Thus

$$\int_\omega X^*(\mathbf{k}_i,\mathbf{r})X(\mathbf{k}_j,\mathbf{r})d\tau=\cos(\mathbf{K}_j-\mathbf{K}_i)\cdot\mathfrak{R}\left[\delta_{ij}-\frac{2}{\Omega}\sum_s \mu_{s,a}^*(\mathbf{k}_i)\mu_{s,a}(\mathbf{k}_j)\right]. \quad (70)$$

Defining

$$\mathfrak{B}_{ij}^q(\mathbf{k})\equiv\mathfrak{A}_q^*(\mathbf{k}_i)\mathfrak{A}_q(\mathbf{k}_j)\cos(\mathbf{K}_j-\mathbf{K}_i)\cdot\mathfrak{R}, \quad (71)$$

we have

$$I_q(\mathbf{k})=\sum_{ij}\mathfrak{B}_{ij}^q(\mathbf{k})\left[\delta_{ij}-\frac{2}{\Omega}\sum_s \mu_{s,a}^*(\mathbf{k}_i)\mu_{s,a}(\mathbf{k}_j)\right]. \quad (72)$$

We may write

$$\sum_n V_a'(\mathbf{r}-\mathbf{R}_n)-\sum_n{}' V_a'(\mathbf{R}_n-\mathbf{R}_m)=\sum_i \cos\mathbf{K}_i\cdot\mathfrak{R}\,\mathfrak{V}_a'(\mathbf{K}_i)e^{-i\mathbf{K}_i\cdot\mathbf{r}}, \quad (73)$$

where

$$\mathfrak{V}_a'(\mathbf{K}_i)\equiv\frac{2}{\Omega}\int_\omega e^{i\mathbf{K}_i\cdot\mathbf{r}}V_a'(\mathbf{r})d\tau-\delta_{i0}\sum_n{}' V_a'(\mathbf{R}_n-\mathbf{R}_m). \quad (74)$$

Thus

$$A_{1q}(\mathbf{k})=\sum_{ij}\mathfrak{B}_{ij}^q(\mathbf{k})\mathfrak{V}_a'(\mathbf{K}_j-\mathbf{K}_i). \quad (75)$$

Similarly, if we define

$$\mathfrak{V}_{ba}(\mathbf{K}_i)\equiv\frac{2}{\Omega}\int_\omega e^{i\mathbf{K}_i\cdot\mathbf{r}}V_{ba}(\mathbf{r})d\tau-\delta_{i0}\sum_n{}' V_{ba}(\mathbf{R}_n-\mathbf{R}_m), \quad (76)$$

then

$$B_{1q}(\mathbf{k})=\sum_{ij}\mathfrak{B}_{ij}^q(\mathbf{k})\mathfrak{V}_{ba}(\mathbf{K}_j-\mathbf{K}_i). \quad (77)$$

B_2 is given by

$$B_{2q}(\mathbf{k})=\frac{2}{\Omega}\sum_{ij}\mathfrak{B}_{ij}^q(\mathbf{k})\sum_s\{[E_q(\mathbf{k})-E_{s,b}]\mu_{s,b}^*(\mathbf{k}_i)\mu_{s,b}(\mathbf{k}_j)-[E_q(\mathbf{k})-E_{s,a}]\mu_{s,a}^*(\mathbf{k}_i)\mu_{s,a}(\mathbf{k}_j)\}, \quad (78)$$

where $\mu_{s,b}$ is defined in a manner analogous to Eq. (66).

In order to determine A_2 we must define

$$\mu_{s,a}'(\mathbf{k}_i)\equiv\int_\infty e^{i\mathbf{k}_i\cdot\mathbf{r}}\varphi_{s,a}'^*(\mathbf{k})d\tau. \tag{79}$$

Thus

$$A_{2q}(\mathbf{k})=\frac{2}{\Omega}\sum_{ij}\mathcal{B}_{ij}{}^q(\mathbf{k})\sum_s\{[3E_q(\mathbf{k})-5E_{s,a}]\times\mu_{s,a}{}^*(\mathbf{k}_i)\mu_{s,a}(\mathbf{k}_j)+[E_q(\mathbf{k})-E_{s,a}]\times[\mu_{s,a}{}^*(\mathbf{k}_i)\mu_{s,a}'(\mathbf{k}_j)+\mu_{s,a}'^*(\mathbf{k}_i)\mu_{s,a}(\mathbf{k}_j)]\}. \tag{80}$$

The divergence theorem of vector analysis tells us that

$$\int_\infty\nabla\cdot[\mathbf{r}e^{i\mathbf{k}_i\cdot\mathbf{r}}\varphi_{s,a}{}^*(\mathbf{r})]d\tau=0,$$

so that

$$\mu_{s,a}'(\mathbf{k}_i)=-[3+\mathbf{k}_i\cdot\nabla_{k_i}]\mu_{s,a}(\mathbf{k}_i), \tag{81}$$

and therefore

$$A_{2q}(\mathbf{k})=-\frac{2}{\Omega}\sum_{ij}\mathcal{B}_{ij}{}^q(\mathbf{k})\sum_s\{2E_q(\mathbf{k})+[E_q(\mathbf{k})-E_{s,a}]\times[\mathbf{k}_i\cdot\nabla_{k_i}+\mathbf{k}_j\cdot\nabla_{k_j}+1]\}\mu_{s,a}{}^*(\mathbf{k}_i)\mu_{s,a}(\mathbf{k}_j). \tag{82}$$

Consistent with Eq. (27), we may define the *real* quantity $Q_{n,l}(r)$ such that

$$\varphi_{nlm}(\mathbf{r})=i^lr^{-1}Q_{nl}(r)\mathcal{Y}_l{}^m(\theta,\varphi), \tag{83}$$

where

$$\int_0^\infty Q_{nl}(r)Q_{n'l}(r)dr=\delta_{nn'}, \tag{84}$$

and

$$\int\mathcal{Y}_l{}^m(\theta,\varphi)^*\mathcal{Y}_{l'}{}^{m'}(\theta,\varphi)d\omega=\delta_{ll'}\delta_{mm'}. \tag{85}$$

(Here $\int d\omega$ denotes a surface integral over the unit sphere.) $\mathcal{Y}_l{}^m$ is defined as

$$\mathcal{Y}_l{}^m(\theta,\varphi)=\left[\left(\frac{2l+1}{4\pi}\right)\frac{(l-|m|)!}{(l+|m|)!}\right]^{\frac{1}{2}}P_l{}^{|m|}(\cos\theta)e^{im\varphi}. \tag{86}$$

Since

$$e^{i\mathbf{k}_i\cdot\mathbf{r}}=4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{+l}i^lj_l(k_ir)\mathcal{Y}_l{}^m(\theta,\varphi)\mathcal{Y}_l{}^m(\theta_i,\varphi_i)^*, \tag{87}$$

(where θ_i and φ_i give the orientation of $\mathbf{k}_i$), we have

$$\mu_{nlm}(\mathbf{k}_i)=4\pi g_{nl}(k_i)\mathcal{Y}_l{}^m(\theta_i,\varphi_i)^*, \tag{88}$$

where

$$g_{nl}(k_i)\equiv\int_0^\infty rQ_{nl}(r)j_l(k_ir)dr. \tag{89}$$

(j_l is the spherical Bessel function of order l.) Let θ_{ij} be the angle between $\mathbf{k}_i$ and $\mathbf{k}_j$. Unsöld's addition theorem states that

$$\sum_{m=-l}^{+l}\mathcal{Y}_l{}^m(\theta_i,\varphi_i)\mathcal{Y}_l{}^m(\theta_j,\varphi_j)^*=(4\pi)^{-1}(2l+1)P_l(\cos\theta_{ij}). \tag{90}$$

Thus we obtain finally

$$I_q(\mathbf{k})=\sum_{ij}\mathcal{B}_{ij}{}^q(\mathbf{k})\left[\delta_{ij}-\frac{8\pi}{\Omega}\sum_l(2l+1)P_l(\cos\theta_{ij})\times\sum_n g_{nla}(k_i)g_{nla}(k_j)\right], \tag{91}$$

$$A_{2q}(\mathbf{k})=-\frac{8\pi}{\Omega}\sum_{ij}\mathcal{B}_{ij}{}^q(\mathbf{k})\sum_{nl}(2l+1)P_l(\cos\theta_{ij})\times\left\{2E_q(\mathbf{k})+[E_q(\mathbf{k})-E_{nla}]\times\left[k_i\frac{\partial}{\partial k_i}+k_j\frac{\partial}{\partial k_j}+1\right]\right\}g_{nla}(k_i)g_{nla}(k_j), \tag{92}$$

$$B_{2q}(\mathbf{k})=\frac{8\pi}{\Omega}\sum_{ij}\mathcal{B}_{ij}{}^q(\mathbf{k})\sum_{nl}(2l+1)P_l(\cos\theta_{ij})\times\{[E_q(\mathbf{k})-E_{nlb}]g_{nlb}(k_i)g_{nlb}(k_j)-[E_q(\mathbf{k})-E_{nla}]g_{nla}(k_i)g_{nla}(k_j)\}. \tag{93}$$

In a specific application, Eqs. (75), (77), (91), (92), and (93) are what need to be evaluated numerically. These equations look considerably more formidable than they really are. Actually they are particularly suitable for numerical computation. The g_{nl}'s which are needed in these equations can be obtained through Eq. (89) from the radial ion-core wave functions $r^{-1}Q_{nl}$ appropriate to isolated atoms in question. These Q_{nl}'s are available in the literature for many atoms, as are the corresponding E_{nl}'s. As can be seen from Eq. (41), these E_{nl}'s are the ion-core energy levels appropriate to the *isolated* atoms in question. (E_{nl} and Q_{nl} for the ion-core levels are approximately independent of the assumed degree of ionization of the isolated atom.) For V_a and V_b, needed in calculating $\mathcal{U}_a'$ and $\mathcal{U}_{bb}$, it is probably satisfactory to use the spherically symmetrical atomic potentials appropriate to the two isolated atoms.

It is a pleasure for the writer to acknowledge his indebtedness to Dr. F. Herman and Dr. D. O. North for many stimulating discussions.

Uniform Strains and Deformation Potentials

R. H. PARMENTER
RCA Laboratories, Radio Corporation of America, Princeton, New Jersey
(Received May 31, 1955)

A perturbation method of calculating the energy levels of a crystal modified by an arbitrary uniform strain is studied. A modified form of first-order perturbation theory is used, a form particularly appropriate to cases where the unperturbed wave functions are expanded in terms of orthogonalized plane waves. Not only are the conventional uniform strains considered but also the special types of deformations whereby the relative positions of the atoms in a unit cell are changed without distorting the Bravais lattice of the crystal. Thus the results of this paper may be applied to the deformation-potential theory of mobility with regard to both acoustic and optical lattice vibrations.

I. INTRODUCTION

IN a previous paper the effect of a uniform dilation upon the one-electron energy bands of a monatomic crystal was determined.[1] In this paper we will generalize this to the case of an arbitrary uniform strain. We will use the same method of calculation used in the previous paper; namely, a modified form of first-order perturbation theory which will be discussed in detail in the following section. The quantities to be determined in the present study may be referred to as "deformation potentials," a deformation potential for a given energy level and for a given strain being defined as the difference in energy between the energy level in the strained and unstrained crystals. The deformation potential will, of course, be a linear function of the amount of strain. The concept of a deformation potential was first introduced by Bardeen and Shockley[2] in connection with the study of the scattering of electrons by lattice vibrations in crystals. They demonstrated the fact that a *long-wavelength* lattice vibration has the same effect locally upon an electron of the crystal as does the locally equivalent uniform strain. By assuming that this is also true to a good approximation in the case of short-wavelength lattice vibrations, these authors were able to compute the electron mobility from a knowledge of the deformation potentials. Associated with each sinusoidal lattice vibration, there may be considered to be a sinusoidally varying deformation potential which will scatter electrons in the crystal, the dependence upon position in the crystal of the deformation potential resulting from the dependence upon position of the local strain resulting from the lattice vibration. The long-wavelength *acoustic* lattice vibrations give rise to conventional strains in the crystal; the long-wavelength *optical* lattice vibrations do not give rise to any conventional strain, however, but rather to a special type of deformation whereby the relative positions of the atoms in a unit cell are changed without distorting the Bravais lattice of the crystal. (We restrict ourselves to long wavelengths in order to maintain the condition of local uniformity of strain in the crystal.) For obvious reasons we will refer to such a deformation as an *optical* strain. When we wish to distinguish between the two types, we will refer to the conventional strain as an *acoustic* strain. In Sec. II, we will work out the deformation potentials for acoustic strains, and in Sec. III we will do the same for optical strains. In Sec. IV, we will show how the modified first-order perturbation approach used in this paper is particularly appropriate for a crystal whose wave functions and energy levels have been determined by the orthogonalized plane wave method.[3] As an example we will consider in detail a crystal of the diamond structure. This is a particularly appropriate case, since OPW (orthogonalized plane wave) solutions have been obtained for the three most important crystals of this type, namely, diamond,[4] silicon,[5] and germanium.[6]

Since the crystal potential in Schrödinger's equation may be chosen arbitrarily to the extent of an additive constant, it follows that the one-electron energy levels are arbitrary to the same extent. From the definition of a deformation potential given in the preceding paragraph, it now follows that the deformation potential is also unspecified to the extent of an arbitrary constant (the same constant for all deformation potentials associated with a given strain). In the case of a *uniformly* strained crystal, this is of no consequence, since the only quantities of real physical significance are the *rates of change relative to each other* of the various deformation potentials with strain, these quantities being completely specified. On the other hand, in the case of a *nonuniformly* strained crystal (resulting from the presence of a lattice vibration), the additive constant in the deformation potential does have physical significance and is connected with an additional electrostatic potential (varying with position) which may be set up as a result of the *nonuniformity* of the strain. In their paper, Bardeen and Shockley[2] argue that this additional electrostatic potential is negligible. Although

[1] R. H. Parmenter, preceding paper [Phys. Rev. **99**, 1739 (1955)].
[2] J. Bardeen and W. Shockley, Phys. Rev. **80**, 72 (1950).
[3] C. Herring, Phys. Rev. **57**, 1169 (1940).
[4] F. Herman, Phys. Rev. **88**, 1210 (1952); **93**, 1214 (1954).
[5] T. O. Woodruff, dissertation, California Institute of Technology, 1955 (unpublished); Phys. Rev. **98**, 1741 (1955).
[6] F. Herman and J. Calloway, Phys. Rev. **89**, 518 (1953); F. Herman, Physica **20**, 801 (1954).

this question is perhaps worthy of further study, we shall in this paper assume the correctness of their conclusions. The assumption that this additional electrostatic potential vanishes implies the specification that the crystal potential in the interior of an ion-core of the crystal is unaffected by straining the crystal. Such a specification serves to define uniquely the deformation potentials. This specification will be followed in the succeeding sections of this paper.

The primary purpose of this paper is to develop a method of studying deformation potentials suitable for detailed numerical computations on specific systems. As with the previous paper devoted to crystals modified by alloying or by pressure,[1] the present study arose as a part of a program of computation of various physical properties of the germanium crystal, a program making use of the availability of wave functions and energy levels for germanium.[6] No attempt will be made in this paper to consider further the application of deformation potentials to mobility theory.

II. ACOUSTIC STRAINS

Consider an unstrained monatomic crystal. The one-electron crystal potential may be written in the form

$$\sum_n V(\mathbf{r}-\mathbf{R}_n)-\mathbf{V}(0), \tag{1}$$

where $V(\mathbf{r})$ is an atomic-like potential, the vectors $\mathbf{R}_n$ denote the positions of the atoms of the crystal, and $\mathbf{V}(0)$ is a constant which will be specified presently. The Schrödinger equation for this problem is

$$H\Psi_q(\mathbf{k},\mathbf{r})=E_q(\mathbf{k})\Psi_q(\mathbf{k},\mathbf{r}), \tag{2}$$

where the Hamiltonian is given by

$$H=-\nabla^2+\sum_n V(\mathbf{r}-\mathbf{R}_n)-\mathbf{V}(0). \tag{3}$$

(We are using atomic units.) Here q denotes the particular energy band under consideration, and $\mathbf{k}$ determines the translational properties of the wave function, i.e.,

$$\Psi_q(\mathbf{k},\ \mathbf{r}+\mathbf{R}_n')=\exp(i\mathbf{k}\cdot\mathbf{R}_n')\Psi_q(\mathbf{k},\mathbf{r}), \tag{4}$$

where the $\mathbf{R}_n'$ denote the positions of the centers of the unit cells of the crystal. (The $\mathbf{R}_n'$ will not coincide with the $\mathbf{R}_n$ when there is more than one atom per unit cell.) We wish to consider the effects of a uniform (acoustic) strain[7] upon the crystal, whereupon the positions of the atoms become $\mathfrak{a}\cdot\mathbf{R}_n$ where[8]

$$\mathfrak{a}=\mathfrak{1}+\mathfrak{e}. \tag{5}$$

[7] For brevity, the word *acoustic* will be dropped throughout the remainder of this section but is to be understood in all references to strain.

[8] We use the above symbolic tensor notation (boldface German script) in order to minimize the number of subscripts in this paper. $\mathfrak{a}$ may be considered to be the 3×3 matrix whose elements are a_{ij}. The dot product of two tensors, $\mathfrak{a}\cdot\mathfrak{e}$, is itself a tensor and may be considered to be the matrix whose elements are $\sum_p a_{ip}e_{pj}$ i.e., the product of the two matrices $\mathfrak{a}$ and $\mathfrak{e}$. The double dot product of two tensors, $\mathfrak{a}:\mathfrak{e}$ is the scalar defined as $\sum_{ij} a_{ij}e_{ji}$, which may be considered to be the trace of the matrix $\mathfrak{a}\cdot\mathfrak{e}$. The dot product of a tensor and a vector, $\mathfrak{a}\cdot\mathbf{c}$, is that vector whose components are $\sum_p a_{ip}c_p$. The dyadic tensor $\mathbf{cd}$ is the direct product of the two vectors $\mathbf{c}$ and $\mathbf{d}$ and may be considered to be a matrix whose elements are c_id_j. The transpose of $\mathbf{cd}$, namely that matrix obtained by interchanging rows and columns of the $\mathbf{cd}$ matrix, will sometimes be written as $\mathbf{cd}^\dagger$.

$\mathfrak{1}$ is the unit tensor (whose elements are δ_{ij}), and $\mathfrak{e}$ is the symmetric strain tensor. We shall assume that the crystal potential appropriate to the strained crystal is

$$\sum_n V(\mathbf{r}-\mathfrak{a}\cdot\mathbf{R}_n)-\mathbf{V}(\mathfrak{e}), \tag{6}$$

where the constant $\mathbf{V}(\mathfrak{e})$ is given by

$$\mathbf{V}(\mathfrak{e})=\sum_n{}' V(\mathfrak{a}\cdot[\mathbf{R}_n-\mathbf{R}_m]). \tag{7}$$

The prime over the summation sign denotes that the term for $\mathbf{R}_n=\mathbf{R}_m$ is omitted. We shall restrict ourselves to crystals having sufficient symmetry for $\mathbf{V}(\mathfrak{e})$ to be independent of the index m. The constant $\mathbf{V}(0)$ occurring in Eq. (1) is now defined as $\mathbf{V}(\mathfrak{e})$ for $\mathfrak{e}=0$. Because of these additive constants in Eqs. (1) and (6), it can be seen that, in the neighborhood of any ion-core of the crystal, the crystal potential can be approximated by the atomic-like potential centered on this ion core, both for the strained and the unstrained lattices. Thus the crystal potential in the neighborhood of an ion-core is unaffected by straining the crystal, as was specified in the previous section. Consistent with such a requirement, the only simple assumption concerning the variation of the crystal potential with strain seems to be that expressed by Eq. (6). The Schrödinger equation for the strained crystal is

$$H'\Psi_q'(\mathbf{k},\mathbf{r})=E_q'(\mathbf{k})\Psi_q'(\mathbf{k},\mathbf{r}), \tag{8}$$

where the Hamiltonian is given by

$$H'=-\nabla^2+\sum_n V(\mathbf{r}-\mathfrak{a}\cdot\mathbf{R}_n)-\mathbf{V}(\mathfrak{e}). \tag{9}$$

In analogy with Eq. (4), we have

$$\Psi_q'(\mathbf{k},\ \mathbf{r}+\mathfrak{a}\cdot\mathbf{R}_n')=\exp[i\mathbf{k}\cdot(\mathfrak{a}\cdot\mathbf{R}_n')]\Psi_q'(\mathbf{k},\mathbf{r}). \tag{10}$$

We define $\mathfrak{a}^{-1}$ such that

$$\mathfrak{a}^{-1}\cdot\mathfrak{a}\equiv\mathfrak{1}, \tag{11}$$

and define

$$\Upsilon_q'(\mathbf{k},\mathbf{r})\equiv\Psi_q(\mathfrak{a}\cdot\mathbf{k},\ \mathfrak{a}^{-1}\cdot\mathbf{r}). \tag{12}$$

Making use of Eq. (4) and the fact that

$$[\mathfrak{a}\cdot\mathbf{k}]\cdot[\mathfrak{a}^{-1}\cdot(\mathfrak{a}\cdot\mathbf{R}_n')]=\mathbf{k}\cdot(\mathfrak{a}\cdot\mathbf{R}_n'),$$

we have

$$\Upsilon_q'(\mathbf{k},\ \mathbf{r}+\mathfrak{a}\cdot\mathbf{R}_n')=\exp[i\mathbf{k}\cdot(\mathfrak{a}\cdot\mathbf{R}_n')]\Upsilon_q'(\mathbf{k},\mathbf{r}), \tag{13}$$

so that Υ' has the correct translational periodicity properties for the strained lattice. This suggests that we take the Υ''s as the unperturbed wave functions in a first-order degenerate perturbation calculation, a procedure which is quite straightforward once the wave functions Ψ appropriate to the unstrained crystal have been obtained. There is, however, one basic objection to the use of Υ''s as the unperturbed wave functions. We note that Eq. (12) represents a scaling

of the wave function appropriate to the unstrained crystal, this scaling to be done in both r- and k-space. In reality, however, when a crystal is strained, the change in the wave function cannot be represented by scaling alone (at least in r-space). In fact, in the immediate vicinity of a particular ion core of the crystal, the wave function, rather than being scaled, probably remains unchanged (aside from a phase factor) since as a first approximation the crystal potential remains unchanged.

The above objection can be largely overcome by replacing Ψ' as our unperturbed wave function for the strained crystal by another wave function, the latter having been made orthogonal to the ion-core wave functions appropriate to the strained crystal. We now give the details of this procedure. First we consider the unstrained crystal. We designate by $\Phi_s(\mathbf{k},\mathbf{r})$ any ion-core Bloch wave function appropriate to this case and by $\Psi_q(\mathbf{k},\mathbf{r})$ any of the other Bloch wave functions appropriate to this case. Since the Ψ's and the Φ's are solutions to the same Schrödinger equation, the Ψ's must be orthogonal to the Φ's. In analogy with Herring's method of defining an orthogonalized plane wave,[3] we consider a function $\Gamma_q(\mathbf{k},\mathbf{r})$ such that

$$\Psi_q(\mathbf{k},\mathbf{r})=\Gamma_q(\mathbf{k},\mathbf{r})-\sum_s \Phi_s(\mathbf{k},\mathbf{r})\int_\infty \Phi_s^*(\mathbf{k},\mathbf{r})\Gamma_q(\mathbf{k},\mathbf{r})d\tau. \quad (14)$$

($\int d\tau$ denotes an integration over the entire volume of the crystal.) Assuming the orthogonality of the Φ's,

$$\int_\infty \Phi_s^*(\mathbf{k},\mathbf{r})\Phi_t(\mathbf{k},\mathbf{r})d\tau=\delta_{s,t}, \quad (15)$$

we see that Eq. (14) merely expresses the fact that the Ψ's are orthogonal to the Φ's, i.e.,

$$\int_\infty \Phi_s^*(\mathbf{k},\mathbf{r})\Psi_q(\mathbf{k},\mathbf{r})d\tau=0. \quad (16)$$

In the case of a monatomic host crystal where there is more than one atom per unit cell of the crystal, the index s appearing in the last three equations really stands for two indices, s and σ, where s denotes the type of atomic orbital from which the Bloch function is composed, while σ denotes with which atom in the unit cell the Bloch function is associated; i.e.,

$$\Phi_{s,\sigma}(\mathbf{k},\mathbf{r})=N^{-\frac{1}{2}}\sum_n \exp[i\mathbf{k}\cdot(\mathbf{R}_n'+\mathfrak{R}_\sigma)]\times\varphi_s(\mathbf{r}-\mathbf{R}_n'-\mathfrak{R}_\sigma). \quad (17)$$

The vector $\mathfrak{R}_\sigma$ denotes the position of the σth atom of the unit cell with respect to the center of the unit cell. As before, the Bravais lattice vectors $\mathbf{R}_n'$ denote the positions of the centers of the unit cells, while the atomic positions, previously denoted by $\mathbf{R}_n$, can be written as $\mathbf{R}_n'+\mathfrak{R}_\sigma$. N denotes the number of unit cells in the complete crystal. Let us take the ion-core orbitals on a given atomic site to be orthonormal, and assume that the atomic orbitals on neighboring atomic sites do not overlap, i.e.,

$$\int_\infty \varphi_s^*(\mathbf{r}-\mathbf{R}_m'-\mathfrak{R}_\sigma)\varphi_t(\mathbf{r}-\mathbf{R}_n'-\mathfrak{R}_\tau)d\tau=\delta_{s,t}\delta_{m,n}\delta_{\sigma,\tau}. \quad (18)$$

We see that Eq. (15) is consistent with Eqs. (17) and (18). Using the fact that

$$\Gamma_q(\mathbf{k},\mathbf{r}+\mathbf{R}_n')=\exp(i\mathbf{k}\cdot\mathbf{R}_n')\Gamma_q(\mathbf{k},\mathbf{r}), \quad (19)$$

we see that Eq. (14) can be rewritten

$$\Psi_q(\mathbf{k},\mathbf{r})=\Gamma_q(\mathbf{k},\mathbf{r})-\sum_{s,n}\varphi_s(\mathbf{r}-\mathbf{R}_n)\times\int_\infty \varphi_s^*(\mathbf{r}-\mathbf{R}_n)\Gamma_q(\mathbf{k},\mathbf{r})d\tau. \quad (20)$$

By means of Eq. (20), we have broken up the wave function $\Psi_q(\mathbf{k},\mathbf{r})$ into two parts, one part

$$-\sum_{s,n}\varphi_s(\mathbf{r}-\mathbf{R}_n)\int_\infty \varphi_s^*(\mathbf{r}-\mathbf{R}_n)\Gamma_q(\mathbf{k},\mathbf{r})d\tau \quad (21)$$

being appreciable only in the immediate vicinity of each ion core of the crystal and representing the rapid variations of the wave function in these regions, while the other part $\Gamma_q(\mathbf{k},\mathbf{r})$ is more smoothly varying and is of importance throughout all regions of the crystal.

Turning to the strained crystal, we shall assume that the crystal potential is essentially unchanged by the strain over the region of a given ion core of the crystal, so that the ion-core Bloch wave functions are

$$\Phi_{s,\sigma}'(\mathbf{k},\mathbf{r})=N^{-\frac{1}{2}}\sum_n \exp[i\mathbf{k}\cdot\mathfrak{a}\cdot(\mathbf{R}_n'+\mathfrak{R}_\sigma)]\times\varphi_s(\mathbf{r}-\mathbf{R}_n'-\mathfrak{R}_\sigma). \quad (22)$$

In analogy with Eq. (14), we construct an approximate wave function for the strained crystal by writing

$$\mathfrak{S}_q'(\mathbf{k},\mathbf{r})=\mathfrak{G}_q'(\mathbf{k},\mathbf{r})-\sum_{s,\sigma}\Phi_{s,\sigma}'(\mathbf{k},\mathbf{r})\times\int_\infty \Phi_{s,\sigma}'^*(\mathbf{k},\mathbf{r})\mathfrak{G}_q'(\mathbf{k},\mathbf{r})d\tau, \quad (23)$$

where

$$\mathfrak{G}_q'(\mathbf{k},\mathbf{r})=\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathfrak{a}^{-1}\cdot\mathbf{r}). \quad (24)$$

Similar to Eqs. (19) and (20), we have

$$\mathfrak{G}_q'(\mathbf{k},\mathbf{r}+\mathfrak{a}\cdot\mathbf{R}_n')=\exp(i\mathbf{k}\cdot\mathfrak{a}\cdot\mathbf{R}_n')\mathfrak{G}_q'(\mathbf{k},\mathbf{r}), \quad (25)$$

so that

$$\mathfrak{S}_q'(\mathbf{k},\mathbf{r})=\mathfrak{G}_q'(\mathbf{k},\mathbf{r})-\sum_{s,n}\varphi_s(\mathbf{r}-\mathfrak{a}\cdot\mathbf{R}_n)\times\int_\infty \varphi_s^*(\mathbf{r}-\mathfrak{a}\cdot\mathbf{R}_n)\mathfrak{G}_q'(\mathbf{k},\mathbf{r})d\tau. \quad (26)$$

Like Ψ_q, the wave function $\mathfrak{S}_q'$ consists of two parts, the portion $\mathfrak{G}_q'$ having been scaled from the corre-

sponding portion of Ψ_q appropriate to the unstrained crystal. The portion

$$-\sum_{s,n}\varphi_s(\mathbf{r}-\mathfrak{a}\cdot\mathbf{R}_n)\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{a}\cdot\mathbf{R}_n)\mathfrak{G}_q'(\mathbf{k},\mathbf{r})d\tau \quad (27)$$

has not been obtained by scaling (21), however, but by replacing (21) by linear combinations of the ion-core atomic orbitals centered on the atomic sites of the strained crystal, the linear combinations being chosen such that the $\mathfrak{S}_q''$s so obtained are orthogonal to the $\Phi_{s,\sigma}''$s, i.e.,

$$\int_\infty\Phi_{s,\sigma}''^*(\mathbf{k},\mathbf{r})\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau=0. \quad (28)$$

We see that $\mathfrak{S}_q'$ has the correct translational periodicity properties for the strained lattice, since

$$\mathfrak{S}_q'(\mathbf{k},\mathbf{r}+\mathfrak{a}\cdot\mathbf{R}_n')=\exp(i\mathbf{k}\cdot\mathfrak{a}\cdot\mathbf{R}_n')\mathfrak{S}_q'(\mathbf{k},\mathbf{r}). \quad (29)$$

The ion-core atomic orbitals used in specifying Φ and Φ' are chosen such that

$$[-\nabla^2+V(\mathbf{r})]\varphi_s(\mathbf{r})=E_s\varphi_s(\mathbf{r}), \quad (30)$$

$V(\mathbf{r})$ being the atomic potential occurring in Eq. (1). This can be done since in the immediate neighborhood of any atomic site of the crystal, the crystal potential may be closely approximated (aside from an additional constant) by the atomic-like potential V centered on that site. We shall assume that V may be approximated by the corresponding isolated-atom potential, so that the φ_s's and the E_s's are those for the isolated atom. Thus, to a good approximation,

$$\begin{aligned}H\varphi_s(\mathbf{r}-\mathbf{R}_n)&=E_s\varphi_s(\mathbf{r}-\mathbf{R}_n),\\ H'\varphi_s(\mathbf{r}-\mathfrak{a}\cdot\mathbf{R}_n)&=E_s\varphi_s(\mathbf{r}-\mathfrak{a}\cdot\mathbf{R}_n).\end{aligned} \quad (31)$$

For a given $\mathfrak{a}\cdot\mathbf{k}$, there will be various sets of degenerate wave functions $\Psi_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})$ associated with the unstrained crystal, the members of a given set all having different point symmetries with respect to their arguments in real space so that they are mutually orthogonal.[9] There will be corresponding sets of functions $\mathfrak{S}_q'(\mathbf{k},\mathbf{r})$, the members of these sets serving as our unperturbed wave functions, the previous objection to the use of the Γ''s as unperturbed wave functions not applying to the use of the $\mathfrak{S}''$s. In other words, for a given $\mathfrak{a}\cdot\mathbf{k}$, we diagonalize $[H'-E'(\mathbf{k})]$ with respect to each set of functions $\mathfrak{S}_q'(\mathbf{k},\mathbf{r})$. Since the Γ_q's have the same point symmetries as do the corresponding Ψ_q's, it follows that the Γ_q's associated with a given set are likewise mutually orthogonal. Since uniform scaling in real space will not affect this orthogonality property, it follows that the corresponding $\mathfrak{G}_q''$s are also mutually orthogonal. Note, however, that the $(\mathfrak{S}_q'-\mathfrak{G}_q')$'s, and thus the $\mathfrak{S}_q''$s themselves, are not mutually orthogonal.

[9] We are disregarding the possibility of *accidental* degeneracies.

It will be necessary to consider matrix elements of unity and of the respective Hamiltonians for both the unstrained and the strained crystals. We may write

$$\begin{aligned}&\int_\infty\Psi_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})\Psi_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\\ &=\delta_{pq}\int_\infty|\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})|^2d\tau-N\sum_{s,\sigma}\int_\infty\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)\\ &\quad\times\Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau, \quad (32)\end{aligned}$$

$$\begin{aligned}&\int_\infty\mathfrak{S}_p'^*(\mathbf{k},\mathbf{r})\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau\\ &=\delta_{pq}\int_\infty|\mathfrak{G}_q'(\mathbf{k},\mathbf{r})|^2d\tau-N\sum_{s,\sigma}\int_\infty\varphi_s(\mathbf{r}-\mathfrak{a}\cdot\mathfrak{R}_\sigma)\\ &\quad\times\mathfrak{G}_p'^*(\mathbf{k},\mathbf{r})d\tau\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{a}\cdot\mathfrak{R}_\sigma)\mathfrak{G}_q'(\mathbf{k},\mathbf{r})d\tau, \quad (33)\end{aligned}$$

$$\begin{aligned}&\int_\infty\Psi_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})H\Psi_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\\ &=\int_\infty\Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})H\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\\ &\quad-N\sum_{s,\sigma}E_s\int_\infty\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\\ &\quad\times\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau, \quad (34)\end{aligned}$$

$$\begin{aligned}&\int_\infty\mathfrak{S}_p'^*(\mathbf{k},\mathbf{r})H'\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau\\ &=\int_\infty\mathfrak{G}_p'^*(\mathbf{k},\mathbf{r})H'\mathfrak{G}_q'(\mathbf{k},\mathbf{r})d\tau\\ &\quad-N\sum_{s,\sigma}E_s\int_\infty\varphi_s(\mathbf{r}-\mathfrak{a}\cdot\mathfrak{R}_\sigma)\mathfrak{G}_p'^*(\mathbf{k},\mathbf{r})d\tau\\ &\quad\times\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{a}\cdot\mathfrak{R}_\sigma)\mathfrak{G}_q'(\mathbf{k},\mathbf{r})d\tau. \quad (35)\end{aligned}$$

Consider a set of n degenerate wave functions $\Psi_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})$, q ranging over n values. The secular equation we must solve is given by

$$\det\left\{\int_\infty\mathfrak{S}_p'^*(\mathbf{k},\mathbf{r})[H'-E'(\mathbf{k})]\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau\right\}=0. \quad (36)$$

There will, of course, be n roots $E'(\mathbf{k})$. We make use of

$$\int_\infty\Psi_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})[H-E(\mathfrak{a}\cdot\mathbf{k})]\Psi_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau=0. \quad (37)$$

(Here we have temporarily dropped the index q on $E(\mathfrak{a}\cdot\mathbf{k})$ since the energy is the same for all n wave functions.) We now divide each element of the secular determinant appearing in Eq. (36) by the constant

$\det\mathfrak{a}$ and from the result subtract Eq. (37), so that Eq. (36) becomes

$$\det\Big\{(\det\mathfrak{a})^{-1}\int_\infty \mathfrak{S}_p'^*(\mathbf{k},\mathbf{r})[H'-E'(\mathbf{k})]\mathfrak{S}_q'(\mathbf{k},\mathbf{r})d\tau$$

$$-\int_\infty \Psi_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})[H-E(\mathfrak{a}\cdot\mathbf{k})]\Psi_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\Big\}=0. \quad (38)$$

($\det\mathfrak{a}$ designates the determinant whose elements equal those of the tensor $\mathfrak{a}$.) In the integrals pertaining to the strained crystal appearing in Eq. (38) we replace the variable of integration $\mathbf{r}$ by $\mathfrak{a}\cdot\mathbf{r}$ so that ∇ is replaced by $\mathfrak{a}^{-1}\cdot\nabla$ and $d\tau$ by $(\det\mathfrak{a})d\tau$. Making use of Eqs. (3), (7), (9), (24), and (32) through (35), we obtain

$$\det\Big\{\int_\infty \Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})[-(\mathfrak{a}^{-1}\cdot\nabla)^2+\nabla^2$$

$$+\sum_n\{V(\mathfrak{a}\cdot[\mathbf{r}-\mathbf{R}_n])-V(\mathbf{r}-\mathbf{R}_n)\}$$

$$-\sum_n{}'\{V(\mathfrak{a}\cdot[\mathbf{R}_n-\mathbf{R}_m])-V(\mathbf{R}_n-\mathbf{R}_m)\}]\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau$$

$$-N\sum_{s,\sigma}E_s\Big[(\det\mathfrak{a})\int_\infty \varphi_s(\mathfrak{a}\cdot[\mathbf{r}-\mathfrak{R}_\sigma])\Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau$$

$$\times\int_\infty \varphi_s^*(\mathfrak{a}\cdot[\mathbf{r}-\mathfrak{R}_\sigma])\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau$$

$$-\int_\infty \varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau$$

$$\times\int_\infty \varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\Big]$$

$$-[E'(\mathbf{k})-E(\mathfrak{a}\cdot\mathbf{k})]\int_\infty \Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau$$

$$+N\sum_{s,\sigma}\Big[(\det\mathfrak{a})E'(\mathbf{k})\int_\infty \varphi_s(\mathfrak{a}\cdot[\mathbf{r}-\mathfrak{R}_\sigma])$$

$$\times\Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\int_\infty \varphi_s^*(\mathfrak{a}\cdot[\mathbf{r}-\mathfrak{R}_\sigma])\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau$$

$$-E(\mathfrak{a}\cdot\mathbf{k})\int_\infty \varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau$$

$$\times\int_\infty \varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau\Big]\Big\}=0. \quad (39)$$

Since we desire the eigenvalues $E'(\mathbf{k})$ only to the accuracy of terms linear in the strain tensor $\mathfrak{e}$, we may define the tensor $\mathfrak{D}(\mathbf{k})$, which is independent of $\mathfrak{e}$, such that

$$E'(\mathbf{k})=E(\mathbf{k})+\mathfrak{e}:\mathfrak{D}(\mathbf{k}). \quad (40)$$

We substitute (40) into (39) and expand everything as a power series in $\mathfrak{e}$, retaining only the lowest order nonvanishing terms. Thus we expand

$$E(\mathfrak{a}\cdot\mathbf{k})=E(\mathbf{k})+\mathfrak{e}:[\mathbf{k}\mathbf{v}(\mathbf{k})+\mathbf{v}(\mathbf{k})\mathbf{k}], \quad (41)$$

where the velocity $\mathbf{v}$ is given by

$$\mathbf{v}(\mathbf{k})=\tfrac{1}{2}\nabla_{\mathbf{k}}E(\mathbf{k}). \quad (42)$$

Similarly we expand

$$-(\mathfrak{a}^{-1}\cdot\nabla)^2=-\nabla^2+\mathfrak{e}:2\nabla\nabla, \quad (43)$$

and

$$V(\mathfrak{a}\cdot\mathbf{r})=V(\mathbf{r})+\mathfrak{e}:\mathfrak{U}'(\mathbf{r}), \quad (44)$$

where

$$\mathfrak{U}'(\mathbf{r})=\mathbf{r}\nabla V(\mathbf{r}). \quad (45)$$

We shall henceforth assume that $V(\mathbf{r})$ is a function of the *magnitude* of $\mathbf{r}$ alone, so that

$$\mathfrak{U}'(\mathbf{r})=\mathbf{r}\mathbf{r}\left(\frac{1}{r}\frac{dV}{dr}\right)$$

$$=-\nabla\nabla\int_r^\infty rV(r)dr-\mathfrak{1}V(\mathbf{r}). \quad (46)$$

We also expand

$$\det\mathfrak{a}=1+\mathfrak{e}:\mathfrak{1}, \quad (47)$$

and

$$\varphi_s(\mathfrak{a}\cdot\mathbf{r})=\varphi_s(\mathbf{r})+\mathfrak{e}:\tfrac{1}{2}[\mathbf{r}\nabla\varphi_s(\mathbf{r})+\{\nabla\varphi_s(\mathbf{r})\}\mathbf{r}]. \quad (48)$$

Substituting Eqs. (40), (41), (43), (44), (47), and (48) into Eq. (39), making use of the facts that

$$\int_\infty [\Psi_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})-\Gamma_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})][H-E(\mathfrak{a}\cdot\mathbf{k})]$$

$$\times[\Psi_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})-\Gamma_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})]d\tau=0, \quad (49)$$

and

$$\int_\infty \Psi_p^*(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})\Psi_q(\mathfrak{a}\cdot\mathbf{k},\mathbf{r})d\tau=0, \quad (50)$$

when $p\neq q$, and dividing the pqth element of the secular determinant by $[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{\frac{1}{2}}$ where

$$I_q(\mathbf{k})=\int_\infty|\Gamma_q(\mathbf{k})|^2d\tau$$

$$-N\sum_{s,\sigma}\left|\int_\infty \varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2, \quad (51)$$

we finally obtain

$$\det\mathfrak{e}:\{\mathfrak{A}_{pq}(\mathbf{k})+\mathfrak{B}_{pq}(\mathbf{k})+\mathfrak{C}_{pq}(\mathbf{k})$$

$$-[\mathfrak{D}(\mathbf{k})-\mathbf{k}\mathbf{v}_q(\mathbf{k})-\mathbf{v}_q(\mathbf{k})\mathbf{k}]\delta_{pq}\}=0, \quad (52)$$

where

$$\mathfrak{A}_{pq}(\mathbf{k})=[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\int_\infty \Gamma_p^*(\mathbf{k},\mathbf{r})2\nabla\nabla\Gamma_q(\mathbf{k},\mathbf{r})d\tau, \quad (53)$$

$$\mathfrak{B}_{pq}(\mathbf{k})=[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-1}\int_\infty \Gamma_p^*(\mathbf{k},\mathbf{r})\Gamma_q(\mathbf{k},\mathbf{r})\times[\sum_n \mathfrak{U}'(\mathbf{r}-\mathbf{R}_n)-\sum_n{}'\mathfrak{U}'(\mathbf{R}_n-\mathbf{R}_m)]d\tau, \quad (54)$$

$$\mathfrak{G}_{pq}(\mathbf{k})=[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}N\sum_{s,\sigma}[E(\mathbf{k})-E_s]\times\left[\delta_{pq}1\left|\int_\infty \varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\right|^2\right.$$
$$+\tfrac{1}{2}\int_\infty \Gamma_p^*(\mathbf{k},\mathbf{r})\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)d\tau\times\int_\infty \Gamma_q(\mathbf{k},\mathbf{r})\{(\mathbf{r}-\mathfrak{R}_\sigma)\nabla\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)$$
$$+(\nabla\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma))(\mathbf{r}-\mathfrak{R}_\sigma)\}d\tau$$
$$+\tfrac{1}{2}\int_\infty \Gamma_q(\mathbf{k},\mathbf{r})\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)d\tau\times\int_\infty \Gamma_p^*(\mathbf{k},\mathbf{r})\{(\mathbf{r}-\mathfrak{R}_\sigma)\nabla\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)$$
$$\left.+(\nabla\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma))(\mathbf{r}-\mathfrak{R}_\sigma)\}d\tau\right]. \quad (55)$$

Equation (52) immediately may be rewritten in the form

$$\det\{\mathfrak{A}_{pq}(\mathbf{k})+\mathfrak{B}_{pq}(\mathbf{k})+\mathfrak{G}_{pq}(\mathbf{k})-[\mathfrak{D}(\mathbf{k})-\mathbf{k}\mathbf{v}_q(\mathbf{k})-\mathbf{v}_q(\mathbf{k})\mathbf{k}]\delta_{pq}\}=0. \quad (56)$$

Equation (56) symbolically represents six sets of secular equations, each set individually determining one of the six components of the desired symmetric tensor $\mathfrak{D}(\mathbf{k})$ for all values of $\mathbf{k}$ allowed by the periodic boundary conditions. The fact that $\mathfrak{D}$ is symmetric follows from the fact that each of the four tensors $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{G}$ and $(\mathbf{kv}+\mathbf{vk})$ are obviously symmetric.

Equation (56) is the general solution to the problem of uniform (acoustic) strains. Note that for cubic crystals the four above-mentioned tensors, and thus also $\mathfrak{D}$, all have cubic symmetry in k-space. Thus for a given tensor any diagonal element may be obtained from any other diagonal element by an appropriate rotation in k-space about one of the cube axes. The nondiagonal elements are similarly related. Physically this means that for cubic crystals it is sufficient to study (for all values of $\mathbf{k}$) the effects of a pure tensile strain along one of cube axes and the effects of a pure shear strain in one of the cube planes.

An interesting special case concerning the cubic crystal occurs when we examine a *nondegenerate* band for $\mathbf{k}$ lying along a symmetry axis. By a proper orientation of our coordinate axes (such that one of them lies parallel to the symmetry axis), we can make $\mathfrak{D}$ a diagonal tensor having only three (diagonal) components. In general, each of these three components will be different, but in the case of a three-fold or a four-fold symmetry axis, two of these components will be identical (those two associated with the two coordinate directions perpendicular to the symmetry axis). In this latter case the effect of an arbitrary strain may be determined from a knowledge of the effects of (*a*) a pure dilation and (*b*) a uniaxial shear along the symmetry direction (the latter being defined as a tensile strain of a given value along the axis and tensile strains of minus one-half this value along each of the other two coordinate directions). Such a case is of practical importance since it includes the bottom of the conduction band in both silicon and germanium, the bottom lying along the fourfold (100) axes in silicon and along the threefold (111) axes in germanium. It should be emphasized that the above conclusions, which follow from symmetry considerations, hold only for $\mathbf{k}$ along a symmetry axis and only for a nondegenerate band along this axis.

III. OPTICAL STRAINS

In this section, we wish to consider the effects of a uniform optical strain[10] upon a crystal. As much as possible, the notation developed in the previous section will be used here. Thus the Hamiltonian for the unstrained crystal is given by Eq. (3), while the Hamiltonian for the strained crystal is

$$H''=-\nabla^2+\sum_n V(\mathbf{r}-\mathbf{R}_n'')-\mathbf{V}(\mathbf{e}_\sigma). \quad (57)$$

Here $\mathbf{V}(\mathbf{e}_\sigma)$ is a constant which will be defined presently, and the $\mathbf{R}_n''$ represent the positions of the atoms in the strained crystal. The positions of the unit cells of the crystal, denoted by the $\mathbf{R}_n'$, are unaffected by the strain, but the positions of the various atoms of a unit cell relative to the center of that cell are changed by the strain from $\mathfrak{R}_\sigma$ to

$$\mathfrak{R}_\sigma''=\mathfrak{R}_\sigma+\mathbf{e}_\sigma. \quad (58)$$

The $\mathbf{e}_\sigma$ are infinitesimal vectors which specify the strain. It can be seen that

$$\mathbf{R}_n''=\mathbf{R}_n'+\mathbf{R}_\sigma''=\mathbf{R}_n+\mathbf{e}_{\sigma(n)}. \quad (59)$$

We shall restrict ourselves to crystals having sufficient symmetry for $\mathbf{V}(\mathbf{e}_\sigma)$, defined as

$$\mathbf{V}(\mathbf{e}_\sigma)=\sum_n{}' V(\mathbf{R}_n''-\mathbf{R}_m''), \quad (60)$$

to be independent of the index m, at least in the limit of small $\mathbf{e}_\sigma$. (The prime over the summation indicates that the term $n=m$ is omitted.) This is necessary in order for the crystal potential in the neighborhood of any ion core of the crystal to be unaffected by the strain. Noting that, for the case $\mathbf{e}_\sigma=0$, Eq. (60) becomes the $\mathbf{V}(0)$ appearing in Eq. (3), we see that, in the neighborhood of any ion core of the crystal, the crystal

[10] For brevity, the word *optical* will be dropped throughout the remainder of this section but is to be understood in all references to strain.

potential can be approximated by the atomic-like potential V centered on this ion core, both for the strained and the unstrained crystals. Since the $\mathbf{R}_n'$ are unaffected by the strain, it follows that the translational periodicity properties of the wave functions for both the strained and the unstrained crystals are identical. This fact, plus the fact that here we are dealing with the vectors $\mathbf{e}_\sigma$ rather than the tensor $\mathbf{e}$, makes the treatment of optical strains somewhat simpler than that of acoustic strains.

We shall use the analog of the method used in the previous section in order to construct an approximate wave function for the strained crystal. Thus, in analogy with Eqs. (20) and (26), we take our approximate wave function to be

$$\mathfrak{S}_q''(\mathbf{k},\mathbf{r})=\Gamma_q(\mathbf{k},\mathbf{r})-\sum_{s,n}\varphi_s(\mathbf{r}-\mathbf{R}_n'')\times\int_\infty\varphi_s^*(\mathbf{r}-\mathbf{R}_n'')\Gamma_q(\mathbf{k},\mathbf{r})d\tau. \quad (61)$$

Note that here it is not necessary to scale Γ since the translational periodicities are unaffected by the strain. For a given $\mathbf{k}$ there will be various sets of degenerate wave functions Ψ_q associated with the unstrained crystal, the members of a given set all having different point symmetries with respect to their arguments in real space so that they are mutually orthogonal.[9] Since the Γ_q have the same point symmetries as do the corresponding Ψ_q, it follows that the Γ_q associated with a given set are likewise mutually orthogonal. There will be corresponding sets of functions $\mathfrak{S}_q''$, the members of these sets serving as our unperturbed wave functions. In other words, for a given $\mathbf{k}$, we diagonalize $[H''-E''(\mathbf{k})]$ with respect to each set of functions $\mathfrak{S}_q''$. Note that the $(\mathfrak{S}_q''-\Gamma_q)$, and thus the $\mathfrak{S}_q''$ themselves, are not mutually orthogonal. It will be necessary to consider matrix elements of unity and of the respective Hamiltonians for both the unstrained and the strained crystals. We may write

$$\int_\infty\Psi_p^*(\mathbf{k},\mathbf{r})\Psi_q(\mathbf{k},\mathbf{r})d\tau=\delta_{pq}\int_\infty|\Gamma_q(\mathbf{k},\mathbf{r})|^2d\tau-N\sum_{s,\sigma}\int_\infty\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)\times\Gamma_p^*(\mathbf{k},\mathbf{r})d\tau\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau, \quad (62)$$

$$\int_\infty\mathfrak{S}_p''^*(\mathbf{k},\mathbf{r})\mathfrak{S}_q''(\mathbf{k},\mathbf{r})d\tau=\delta_{pq}\int_\infty|\Gamma_q(\mathbf{k},\mathbf{r})|^2d\tau-N\sum_{s,\sigma}\int_\infty\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma'')\times\Gamma_p^*(\mathbf{k},\mathbf{r})d\tau\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma'')\Gamma_q(\mathbf{k},\mathbf{r})d\tau, \quad (63)$$

$$\int_\infty\Psi_p^*(\mathbf{k},\mathbf{r})H\Psi_q(\mathbf{k},\mathbf{r})d\tau=\int_\infty\Gamma_p^*(\mathbf{k},\mathbf{r})H\Gamma_q(\mathbf{k},\mathbf{r})d\tau-N\sum_{s,\sigma}E_s\int_\infty\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)\times\Gamma_p^*(\mathbf{k},\mathbf{r})d\tau\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau, \quad (64)$$

$$\int_\infty\mathfrak{S}_p''^*(\mathbf{k},\mathbf{r})H''\mathfrak{S}_q''(\mathbf{k},\mathbf{r})d\tau=\int_\infty\Gamma_p^*(\mathbf{k},\mathbf{r})H''\Gamma_q(\mathbf{k},\mathbf{r})d\tau-N\sum_{s,\sigma}E_s\int_\infty\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma'')\Gamma_p^*(\mathbf{k},\mathbf{r})d\tau\times\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma'')\Gamma_q(\mathbf{k},\mathbf{r})d\tau. \quad (65)$$

Equations (62) and (64) can, of course, be obtained directly from Eqs. (32) and (34), respectively, simply by replacing $\mathfrak{a}\cdot\mathbf{k}$ by $\mathbf{k}$. In obtaining Eq. (65) we made use of the fact that

$$H''\varphi_s(\mathbf{r}-\mathbf{R}_n'')=E_s\varphi_s(\mathbf{r}-\mathbf{R}_n'') \quad (66)$$

to a good approximation, in analogy with Eq. (31). Consider a set of n degenerate wave functions $\Psi_q(\mathbf{k},\mathbf{r})$, q ranging over n values. The secular equation we must solve is given by

$$\det\left\{\int_\infty\mathfrak{S}_p''^*(\mathbf{k},\mathbf{r})[H''-E''(\mathbf{k})]\mathfrak{S}_q''(\mathbf{k},\mathbf{r})d\tau\right\}=0. \quad (67)$$

There will, of course, be n roots $E''(\mathbf{k})$. In Eq. (37) we replace $\mathfrak{a}\cdot\mathbf{k}$ by $\mathbf{k}$ and subtract this from Eq. (67), obtaining

$$\det\left\{\int_\infty\mathfrak{S}_p''^*(\mathbf{k},\mathbf{r})[H''-E''(\mathbf{k})]\mathfrak{S}_q''(\mathbf{k},\mathbf{r})d\tau-\int_\infty\Psi_p^*(\mathbf{k},\mathbf{r})[H-E(\mathbf{k})]\Psi_q(\mathbf{k},\mathbf{r})d\tau\right\}=0. \quad (68)$$

Making use of Eqs. (3), (57), (60), and (62) through (65), we find that Eq. (68) becomes

$$\det\left\{\int_\infty\Gamma_p^*(\mathbf{k},\mathbf{r})[\sum_n\{V(\mathbf{r}-\mathbf{R}_n'')-V(\mathbf{r}-\mathbf{R}_n)\}-\sum_n\{V(\mathbf{R}_n''-\mathbf{R}_m'')-V(\mathbf{R}_n-\mathbf{R}_m)\}-\delta_{pq}\{E''(\mathbf{k})-E(\mathbf{k})\}]\Gamma_q(\mathbf{k},\mathbf{r})d\tau+N\sum_{s,\sigma}\Big[\{E''(\mathbf{k})-E_s\}\int_\infty\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma'')\Gamma_p^*(\mathbf{k},\mathbf{r})d\tau\times\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma'')\Gamma_q(\mathbf{k},\mathbf{r})d\tau-\{E(\mathbf{k})-E_s\}\times\int_\infty\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_p^*(\mathbf{k},\mathbf{r})d\tau\times\int_\infty\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)\Gamma_q(\mathbf{k},\mathbf{r})d\tau\Big]\right\}=0. \quad (69)$$

Since we desire the eigenvalues $E''(\mathbf{k})$ only to the accuracy of terms linear in the strain vectors $\mathbf{e}_\sigma$, we can expand everything as a power series in the $\mathbf{e}_\sigma$, retaining only the lowest order nonvanishing terms. Thus we define the vector $\mathbf{F}_\sigma$ such that

$$E''(\mathbf{k})-E(\mathbf{k})=\sum_\sigma \mathbf{e}_\sigma\cdot\mathbf{F}_\sigma(\mathbf{k}). \tag{70}$$

We shall at this point assume that the crystal symmetry is such that

$$\nabla\sum_n V(\mathbf{R}_n'+\mathfrak{R}_\sigma-\mathfrak{R}_\tau)=0, \tag{71}$$

for any two distinct position vectors $\mathfrak{R}_\sigma$ and $\mathfrak{R}_\tau$. Thus it follows that

$$\sum_n\{V(\mathbf{R}_n''-\mathbf{R}_m'')-V(\mathbf{R}_n-\mathbf{R}_m)\}$$
$$=\sum_\sigma(\mathbf{e}_\sigma-\mathbf{e}_\tau)\cdot\nabla\sum_n V(\mathbf{R}_n'+\mathfrak{R}_\sigma-\mathfrak{R}_\tau)=0, \tag{72}$$

$\mathfrak{R}_\tau$ being associated with $\mathbf{R}_m$. Similarly expanding the other terms in Eq. (69), and making use of Eqs. (49) and (50) with $\mathfrak{a}\cdot\mathbf{k}$ replaced by $\mathbf{k}$, we finally obtain

$$\det\sum_\sigma \mathbf{e}_\sigma\cdot[{}_\sigma\mathbf{X}_{pq}(\mathbf{k})+{}_\sigma\mathbf{Y}_{pq}(\mathbf{k})-\delta_{pq}\mathbf{F}_\sigma(\mathbf{k})]=0, \tag{73}$$

where

$${}_\sigma\mathbf{X}_{pq}(\mathbf{k})=[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\int_\infty \Gamma_p^*(\mathbf{k},\mathbf{r})\Gamma_q(\mathbf{k},\mathbf{r})$$
$$\times\sum_n \nabla V(\mathbf{r}-\mathbf{R}_n'-\mathbf{R}_\sigma)d\tau, \tag{74}$$

$${}_\sigma\mathbf{Y}_{pq}(\mathbf{k})=[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}N\sum_s[E(\mathbf{k})-E_s]$$
$$\times\Big[\int_\infty \Gamma_p^*(\mathbf{k},\mathbf{r})\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)d\tau$$
$$\times\int_\infty \Gamma_q\nabla\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)d\tau$$
$$+\int_\infty \Gamma_q(\mathbf{k},\mathbf{r})\varphi_s^*(\mathbf{r}-\mathfrak{R}_\sigma)d\tau$$
$$\times\int_\infty \Gamma_p^*(\mathbf{k},\mathbf{r})\nabla\varphi_s(\mathbf{r}-\mathfrak{R}_\sigma)d\tau\Big], \tag{75}$$

and $I_q(\mathbf{k})$ is given by Eq. (51). Equation (73) may be rewritten in the form

$$\det[{}_\sigma\mathbf{X}_{pq}(\mathbf{k})+{}_\sigma\mathbf{Y}_{pq}(\mathbf{k})-\delta_{pq}\mathbf{F}_\sigma(\mathbf{k})]=0. \tag{76}$$

Equation (76) symbolically represents 3ν sets of equations, ν being the number of atoms per unit cell of the crystal, each set individually determining one of 3ν components $F_{\sigma(\alpha,p,s)}(\mathbf{k})$ for all values of $\mathbf{k}$ allowed by the periodic boundary conditions.

IV. APPLICATION TO THE DIAMOND STRUCTURE

The procedure developed in the preceding sections is particularly suited to an unmodified crystal for which wave functions and energy levels have been calculated by the orthogonalized plane wave method.[3] In this method the wave function is expanded in terms of orthogonalized plane waves $X(\mathbf{k}_i,\mathbf{r})$, $\mathbf{k}_i$ being defined as

$$\mathbf{k}_i=\mathbf{k}+\mathbf{K}_i. \tag{77}$$

$\mathbf{K}_i$ being a vector of the reciprocal lattice such that

$$\exp(i\mathbf{K}_i\cdot\mathbf{R}_n')=1, \tag{78}$$

for all i and n. $X(\mathbf{k}_i,\mathbf{r})$ is defined as

$$X(\mathbf{k}_i,\mathbf{r})=(N\Omega)^{-\frac{1}{2}}e^{i\mathbf{k}_i\cdot\mathbf{r}}-\Omega^{-\frac{1}{2}}\sum_{s,\sigma}\exp(i\mathbf{K}_i\cdot\mathfrak{R}_\sigma)$$
$$\times\mu_s(\mathbf{k}_i)\Phi_{s,\sigma}(\mathbf{k},\mathbf{r}), \tag{79}$$

where

$$\mu_s(\mathbf{k}_i)=N^{-\frac{1}{2}}\exp(-i\mathbf{K}_i\cdot\mathfrak{R}_\sigma)\int_\infty e^{i\mathbf{k}_i\cdot\mathbf{r}}\Phi_{s,\sigma}^*(\mathbf{k},\mathbf{r})d\tau$$
$$=\int_\infty e^{i\mathbf{k}_i\cdot\mathbf{r}}\varphi_s^*(\mathbf{r})d\tau, \tag{80}$$

and where Ω is the volume of the unit cell of the unstrained crystal. We see now that

$$\int_\infty \Phi_{s,\sigma}^*(\mathbf{k},\mathbf{r})X(\mathbf{k}_i,\mathbf{r})d\tau=0. \tag{81}$$

Expanding $\Psi_q(\mathbf{k},\mathbf{r})$ in terms of the X's, we have

$$\Psi_q(\mathbf{k},\mathbf{r})=\sum_i \mathfrak{a}_q(\mathbf{k}_i)X(\mathbf{k}_i,\mathbf{r}). \tag{82}$$

Hence it follows that $\Gamma_q(\mathbf{k},\mathbf{r})$, as defined by Eq. (14), can be written as

$$\Gamma_q(\mathbf{k},\mathbf{r})=(N\Omega)^{-\frac{1}{2}}\sum_i \mathfrak{a}_q(\mathbf{k}_i)e^{i\mathbf{k}_i\cdot\mathbf{r}}. \tag{83}$$

It is because of the simplicity of Eq. (83) that the OPW method is particularly adapted to the results of Secs. II and III.

We shall now specialize to the case where the unstrained crystal has the diamond structure, this being a case of great practical importance. There are now two atoms per unit cell such that $\mathfrak{R}_\tau=\pm\mathfrak{R}$ where $\mathfrak{R}$ points along the (111) direction, and $2\mathfrak{R}$ is the nearest neighbor distance. The optical strain vectors are thus given by $\mathbf{e}_\sigma=\pm\mathbf{e}$ in order that there be no displacement of the center of mass of the crystal during an optical strain. It follows that Eqs. (73) and (76) may be written in the form

$$\det[\mathbf{X}_{pq}(\mathbf{k})+\mathbf{Y}_{pq}(\mathbf{k})-\delta_{pq}\mathbf{F}(\mathbf{k})]=0, \tag{84}$$

where $\mathbf{X}_{pq}$ is the difference of the two ${}_\sigma\mathbf{X}_{pq}$, and similarly for $\mathbf{Y}_{pq}$ and $\mathbf{F}$. Thus, Eq. (70) becomes

$$E''(\mathbf{k})-E(\mathbf{k})=\mathbf{e}\cdot\mathbf{F}(\mathbf{k}). \tag{85}$$

We define

$$\mathfrak{A}_{ij}^{pq}(\mathbf{k})\equiv\mathfrak{a}_p^*(\mathbf{k}_i)\mathfrak{a}_q(\mathbf{k}_j)\cos(\mathbf{K}_j-\mathbf{K}_i)\cdot\mathfrak{R}, \tag{86}$$

$$\mathfrak{C}_{ij}^{pq}(\mathbf{k})\equiv\mathfrak{a}_p^*(\mathbf{k}_i)\mathfrak{a}_q(\mathbf{k}_j)(\mathbf{K}_j-\mathbf{K}_i)\sin(\mathbf{K}_j-\mathbf{K}_i)\cdot\mathfrak{R}. \tag{87}$$

Since we now have

$$\int_\infty X^*(\mathbf{k}_i,\mathbf{r})X(\mathbf{k}_j,\mathbf{r})d\tau$$

$$=\cos(\mathbf{K}_j-\mathbf{K}_i)\cdot\mathfrak{R}\,[\delta_{ij}-(2/\Omega)\textstyle\sum_s \mu_s^*(\mathbf{k}_i)\mu_s(\mathbf{k}_j)], \quad (88)$$

it follows that

$$I_q(\mathbf{k})=\sum_{ij}\mathfrak{G}_{ij}{}^{pq}(\mathbf{k})[\delta_{ij}-(2/\Omega)\sum_s \mu_s^*(\mathbf{k}_i)\mu_s(\mathbf{k}_j)]. \quad (89)$$

Since for the diamond structure

$$\nabla\nabla\sum_n{}'\int_{|\mathbf{R}_n-\mathbf{R}_m|}^\infty rV(r)dr=0, \quad (90)$$

by symmetry, we have from Eqs. (7) and (46)

$$\textstyle\sum_n{}'\,\mathfrak{U}'(\mathbf{R}_n-\mathbf{R}_m)=-\mathbf{1}V(0), \quad (91)$$

so that we may write

$$\textstyle\sum_n \mathfrak{U}'(\mathbf{r}-\mathbf{R}_n)-\sum_n{}'\,\mathfrak{U}'(\mathbf{R}_n-\mathbf{R}_m)$$
$$=\textstyle\sum_i \cos\mathbf{K}_i\cdot\mathfrak{R}\,\mathfrak{W}'(\mathbf{K}_i)e^{-i\mathbf{K}_i\cdot\mathbf{r}}, \quad (92)$$

where

$$\mathfrak{W}'(\mathbf{K}_i)\equiv(2/\Omega)\int_\infty e^{i\mathbf{K}_i\cdot\mathbf{r}}\mathfrak{U}'(\mathbf{r})d\tau+\delta_{i0}\mathbf{1}V(0). \quad (93)$$

With the aid of Eq. (46), we may write

$$\mathfrak{W}'(\mathbf{K}_i)=-\mathbf{K}_i\mathbf{K}_i\mathfrak{M}(\mathbf{K}_i)+\mathbf{1}\mathfrak{N}(\mathbf{K}_i), \quad (94)$$

$\mathfrak{M}$ and $\mathfrak{N}$ being defined as

$$\mathfrak{M}(\mathbf{K}_i)=-(8\pi/\Omega)\int_0^\infty dr r^2\left(\frac{\sin K_i r}{K_i r}\right)\int_r^\infty r'V(r')dr', \quad (95)$$

$$\mathfrak{N}(\mathbf{K}_i)=-(8\pi/\Omega)\int_0^\infty\left(\frac{\sin K_i r}{K_i r}\right)r^2V(r)dr+\delta_{i0}\mathbf{V}(0). \quad (96)$$

Similarly, we have

$$\textstyle\sum_n \nabla\{V(\mathbf{r}-\mathbf{R}_n{}'-\mathfrak{R})-V(\mathbf{r}-\mathbf{R}_n{}'+\mathfrak{R})\}$$
$$=\textstyle\sum_i \sin\mathbf{K}_i\cdot\mathfrak{R}\,\mathbf{K}_i\,\mathfrak{N}(\mathbf{K}_i)e^{i\mathbf{K}_i\cdot\mathbf{r}}, \quad (97)$$

since

$$-(2i/\Omega)\int_\infty e^{i\mathbf{K}_i\cdot\mathbf{r}}\nabla V(\mathbf{r})d\tau=\mathbf{K}_i\mathfrak{N}(\mathbf{K}_i).$$

It now follows that

$$\mathfrak{A}_{pq}(\mathbf{k})=[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\textstyle\sum_i \mathfrak{G}_{ii}{}^{pq}(\mathbf{k})\mathbf{k}_i\mathbf{k}_i, \quad (98)$$

$$\mathfrak{B}_{pq}(\mathbf{k})=[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\sum_{ij}\mathfrak{G}_{ij}{}^{pq}(\mathbf{k})\mathfrak{W}'(\mathbf{K}_j-\mathbf{K}_i), \quad (99)$$

$$\mathbf{X}_{pq}(\mathbf{k})=[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\sum_{ij}\mathfrak{C}_{ij}{}^{pq}(\mathbf{k})\mathfrak{N}(\mathbf{K}_j-\mathbf{K}_i). \quad (100)$$

The divergence theorem of vector analysis tells us that

$$\int_\infty \nabla[\mathbf{r}e^{i\mathbf{k}_i\cdot\mathbf{r}}\varphi_s^*(\mathbf{r})]d\tau=0$$

$$=\int_\infty e^{i\mathbf{k}_i\cdot\mathbf{r}}\mathbf{r}\nabla^\dagger\varphi_s^*(\mathbf{r})d\tau$$
$$+[\mathbf{k}_i\nabla_{k_i}+1]\int_\infty e^{i\mathbf{k}_i\cdot\mathbf{r}}\varphi_s^*(\mathbf{r})d\tau,$$

so that

$$\int_\infty e^{i\mathbf{k}_i\cdot\mathbf{r}}\mathbf{r}\nabla^\dagger\varphi_s^*(\mathbf{r})d\tau=-[\mathbf{1}+\mathbf{k}_i\nabla_{k_i}]\mu_s(\mathbf{k}_i). \quad (101)$$

(See footnote 8 for the meaning of $\mathbf{ab}^\dagger$.) Similarly

$$\int_\infty e^{i\mathbf{k}_i\cdot\mathbf{r}}\mathbf{r}\nabla\varphi_s^*(\mathbf{r})d\tau=-[\mathbf{1}+\mathbf{k}_i\nabla_{k_i}{}^\dagger]\mu_s(\mathbf{k}_i). \quad (102)$$

We now find that

$$\mathfrak{F}_{pq}(\mathbf{k})=-(2/\Omega)[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\sum_{ij}\mathfrak{G}_{ij}{}^{pq}(\mathbf{k})$$
$$\times\sum_s[E(\mathbf{k})-E_s]\{\mathbf{1}(2-\delta_{pq})$$
$$+\tfrac{1}{2}(\mathbf{k}_i\nabla_{k_i}+\mathbf{k}_i\nabla_{k_i}{}^\dagger+\mathbf{k}_j\nabla_{k_j}+\mathbf{k}_j\nabla_{k_j}{}^\dagger)\}$$
$$\times\mu_s^*(\mathbf{k}_i)\mu_s(\mathbf{k}_j). \quad (103)$$

In an analogous fashion, we find that

$$\mathbf{Y}_{pq}(\mathbf{k})=(2/\Omega)[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\sum_{ij}\mathfrak{C}_{ij}{}^{pq}(\mathbf{k})$$
$$\times\sum_s[E(\mathbf{k})-E_s]\mu_s^*(\mathbf{k}_i)\mu_s(\mathbf{k}_j). \quad (104)$$

Consistent with Eq. (18), we may define the *real* quantity $Q_{nl}(r)$ such that

$$\varphi_{nlm}(\mathbf{r})=i^l r^{-1}Q_{nl}(r)\mathcal{Y}_l{}^m(\theta,\varphi), \quad (105)$$

where

$$\int_0^\infty Q_{nl}(r)Q_{n'l}(r)dr=\delta_{nn'}, \quad (106)$$

and

$$\int\mathcal{Y}_l{}^m(\theta,\varphi)^*\mathcal{Y}_{l'}{}^{m'}(\theta,\varphi)d\omega=\delta_{ll'}\delta_{mm'}. \quad (107)$$

(Here $\int d\omega$ denotes a surface integral over the unit sphere.) $\mathcal{Y}_l{}^m$ is defined as

$$\mathcal{Y}_l{}^m(\theta,\varphi)=\left[\left(\frac{2l+1}{4\pi}\right)\frac{(l-|m|)!}{(l+|m|)!}\right]^{\frac{1}{2}}P_l{}^{|m|}(\cos\theta)e^{im\varphi}. \quad (108)$$

Since

$$e^{i\mathbf{k}_i\cdot\mathbf{r}}=4\pi\sum_{l=0}^\infty\sum_{m=-l}^{+l}i^l j_l(k_i r)\mathcal{Y}_l{}^m(\theta,\varphi)\mathcal{Y}_l{}^m(\theta_i,\varphi_i)^*, \quad (109)$$

(where θ_i and φ_i give the orientation of $\mathbf{k}_i$), we have

$$\mu_{nlm}(\mathbf{k}_i)=4\pi g_{nl}(k_i)\mathcal{Y}_l^m(\theta_i,\varphi_i)^*, \tag{110}$$

where

$$g_{nl}(k_i)\equiv\int_0^\infty rQ_{nl}(r)\,j_l(kr)dr. \tag{111}$$

(j_l is the spherical Bessel function of order l.) Let θ_{ij} be the angle between $\mathbf{k}_i$ and $\mathbf{k}_j$. Unsöld's addition theorem states that

$$\sum_{m=-l}^{+l}\mathcal{Y}_i^m(\theta_i,\varphi_i)\mathcal{Y}_j^m(\theta_j,\varphi_j)^*=\left(\frac{2l+1}{4\pi}\right)P_l(\cos\theta_{ij}). \tag{112}$$

Equations (89) and (104) can now be written as

$$I_q(\mathbf{k})=\sum_{ij}\mathfrak{B}_{ij}^{pq}(\mathbf{k})[\delta_{ij}-(8\pi/\Omega)\sum_l(2l+1)\times P_l(\cos\theta_{ij})\sum_n g_{nl}(k_i)g_{nl}(k_j)], \tag{113}$$

$$\mathbf{Y}_{pq}(\mathbf{k})=(8\pi/\Omega)[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\sum_{ij}\mathfrak{C}_{ij}^{pq}(\mathbf{k})\times\sum_l(2l+1)P_l(\cos\theta_{ij})\sum_n[E(\mathbf{k})-E_{nl}]\times g_{nl}(k_i)g_{nl}(k_j). \tag{114}$$

If we define

$$\mathbf{u}=\mathbf{k}_i+\mathbf{k}_j,\quad \mathbf{v}=\mathbf{k}_i-\mathbf{k}_j,$$

then we have

$$k_ik_j\cos\theta_{ij}=\tfrac{1}{4}(u^2-v^2),$$

and

$$(\mathbf{k}_i\nabla_{k_i}+\mathbf{k}_j\nabla_{k_j})=(\mathbf{u}\nabla_u+\mathbf{v}\nabla_v).$$

Thus it follows that

$$(\mathbf{k}_i\nabla_{k_i}+\mathbf{k}_j\nabla_{k_j})k_ik_j\cos\theta_{ij}=\mathbf{k}_i\mathbf{k}_j+\mathbf{k}_j\mathbf{k}_i, \tag{115}$$

so that

$$\begin{aligned}&\tfrac{1}{2}(\mathbf{k}_i\nabla_{k_i}+\mathbf{k}_i\nabla_{k_i}{}^\dagger+\mathbf{k}_j\nabla_{k_j}+\mathbf{k}_j\nabla_{k_j}{}^\dagger)P_l(\cos\theta_{ij})g(k_i)g(k_j)\\&=(\mathbf{k}_i\mathbf{k}_i/k_i^2)[k_ig'(k_i)P_l(\cos\theta_{ij})\\&\quad-\cos\theta_{ij}P_l'(\cos\theta_{ij})g(k_i)]g(k_j)\\&\quad+(\mathbf{k}_j\mathbf{k}_j/k_j^2)[k_jg'(k_j)P_l(\cos\theta_{ij})\\&\quad-\cos\theta_{ij}P_l'(\cos\theta_{ij})g(k_j)]g(k_i)+(k_ik_j)^{-1}\\&\qquad\times[\mathbf{k}_i\mathbf{k}_j+\mathbf{k}_j\mathbf{k}_i]P_l'(\cos\theta_{ij})g(k_i)g(k_j).\end{aligned} \tag{116}$$

Analogously to Eqs. (113) and (114), we may now write Eq. (103) in the form

$$\begin{aligned}\mathfrak{G}_{pq}(\mathbf{k})=&-(8\pi/\Omega)[I_p^*(\mathbf{k})I_q(\mathbf{k})]^{-\frac{1}{2}}\sum_{ij}\mathfrak{B}_{ij}^{pq}(\mathbf{k})\\&\times\sum_{nl}(2l+1)[E(\mathbf{k})-E_{nl}]\{\mathbf{1}(2-\delta_{pq})\\&\times P_l(\cos\theta_{ij})g_{nl}(k_i)g_{nl}(k_j)\\&+(\mathbf{k}_i\mathbf{k}_i/k_i^2)[k_ig_{nl}'(k_i)P_l(\cos\theta_{ij})\\&-\cos\theta_{ij}P_l'(\cos\theta_{ij})g_{nl}(k_i)]g_{nl}(k_j)\\&+(\mathbf{k}_j\mathbf{k}_j/k_j^2)[k_jg_{nl}'(k_j)P_l(\cos\theta_{ij})\\&-\cos\theta_{ij}P_l'(\cos\theta_{ij})g_{nl}(k_j)]g_{nl}(k_i)\\&+(k_ik_j)^{-1}[\mathbf{k}_i\mathbf{k}_j+\mathbf{k}_j\mathbf{k}_i]\\&\times P_l'(\cos\theta_{ij})g_{nl}(k_i)g_{nl}(k_j)\}.\end{aligned} \tag{117}$$

[The primes in Eqs. (116) and (117) denote differentiation with respect to the argument.] In a specific application, Eqs. (98), (99), (100), (114), and (117) are what need to be evaluated numerically. The statements made at the conclusion of reference 1 also apply to these equations.

The writer is indebted to Dr. F. Herman and Dr. D. O. North for encouragement and for helpful discussion.

X-Ray Scattering Effects Due to Localized Static Lattice Defects

CHARLES W. TUCKER, JR., AND PETER SENIO
Knolls Atomic Power Laboratory, General Electric Company, Schenectady, New York*
(Received June 13, 1955)

Theoretical calculations by Huang show that elastic singularities in a crystalline lattice should produce four specific x-ray scattering effects which include diffuse scattering and an artificial temperature factor. All four of the effects are present in neutron-irradiated crystals of boron carbide, diamond, silicon carbide, and magnesium oxide. Thus the model and calculations of Huang are qualitatively confirmed. It is suggested that it may be possible to determine more precisely the nature of the lattice distortions around localized static lattice defects from a detailed study of the diffuse scattering.

INTRODUCTION

THE literature pertaining to the scattering of x-rays by distorted crystals is very large and some of it dates back to the early days of the development of x-ray diffraction theory.[1] However, much of this literature is concerned with distortions and size limitations which upset the coherence of the scattering between the unit cells of which the crystal is composed. This broadens and changes the shape of the coherent scattering peaks, the Bragg reflections, and the problem becomes that of deducing from the alterations of the peak shapes the nature of the distortions and the size distributions present. With the important exception of the effect of temperature on x-ray scattering, the case in which the distortions are due to small localized lattice defects has not been nearly as extensively studied. Since this paper is concerned with localized *static* lattice defects, temperature effects will not be considered, although reference will be made to them. Localized static lattice defects will be interpreted to mean defects which are small in three dimensions, *viz.*, interstitials, vacancies, oversize atoms, or agglomerates of only several atoms at most. Thus scattering due to dislocations, for example, is excluded.

THEORETICAL

Theoretical calculations of the x-ray scattering by a crystalline lattice containing a random distribution of localized static lattice defects are rather few. While we shall consider only the detailed treatment of the problem given by Huang,[2] the literature will be reviewed briefly.

Laue[3] refers to the problem in his treatment of the x-ray scattering by a solid solution in which the component atoms are of different sizes, but does not give detailed calculations. Ekstein[4] and Matsubara[5] examine the case in which the lattice atoms are given small displacements from their sites in a perfect lattice, expressing the scattering by the distorted lattice by the first two terms in a Taylor's series. Ekstein takes as an example of the displacements around a defect the classical elasticity problem of the displacements in a hollow sphere due to a pressure in the internal cavity.[6] Matsubara obtains his displacements by minimizing the strain energy of the lattice as a function of the displacements. For the case of defects producing displacements over rather large distances, Matsubara points out that his treatment becomes equivalent to that of Ekstein. Both authors show that the defects lead to the production of diffuse maxima surrounding the sharp Bragg reflections at the reciprocal lattice points. Other than this specific x-ray effects are not discussed.

Warren, Averbach, and Roberts[7] discuss an atomic size effect in solid solutions due to the difference in size of solvent and solute atoms. Their calculations predict an asymmetry in the diffuse scattering due to the size effect such that there is an increase in scattering on one side of a reciprocal lattice point and a decrease on the other side. The magnitude of this effect depends not only on the difference in size of the solvent and solute atoms, but also on the difference in their atomic scattering powers. Interest in the present paper will be in distortions of more general types for which the difference in scattering power of matrix and defect atoms is either zero or ignored. Therefore, the size effect computed by Warren, Averbach, and Roberts will not be considered further here.

Zachariasen[8] has given a general theory of x-ray scattering by displacement disorders in a crystal. In this theory displacement of the atoms from their normal lattice sites leads to a decrease in the effective scattering power per unit cell and to the production of diffuse scattering surrounding the reciprocal lattice points.

While detailed calculations based on the previously mentioned work, particularly that of Zachariasen,[8] Ekstein,[4] and Matsubara[5] might well be used in de-

* The Knolls Atomic Power Laboratory is operated by the General Electric Company for the U. S. Atomic Energy Commission.

[1] For a recent survey of this field, the monograph by A. J. C. Wilson, *X-ray Optics—The Diffraction of X-rays by Finite and Imperfect Crystals* (Methuen and Company Ltd., London, 1949) is recommended.

[2] K. Huang, Proc. Roy. Soc. (London) **A190**, 102 (1947).

[3] M. v. Laue, *Röntgenstrahl-Interferenzen* (Edwards Brothers, Inc., Ann Arbor, 1943), p. 179.

[4] H. Ekstein, Phys. Rev. **68**, 120 (1945).

[5] T. J. Matsubara, Proc. Phys. Soc. Japan **7**, 270 (1952).

[6] A. E. H. Love, *Mathematical Theory of Elasticity* (Dover Publications, New York, 1944), p. 142.

[7] Warren, Averbach, and Roberts, J. Appl. Phys. **22**, 1493 (1951).

[8] W. H. Zachariasen, *Theory of X-ray Diffraction in Crystals* (John Wiley and Sons, Inc., New York, 1945), Chap. IV.

termining the x-ray effects produced by localized static lattice defects, a rather complete treatment of a fairly general problem has already been given by Huang.[2] Specifically, Huang considered the x-ray scattering by a face-centered cubic lattice containing a random distribution of elastic singularities which produce elastic displacements of the surrounding atoms of the form $u=c/r^2$, where u is the displacement of an atom a distance r from a singularity of "strength" c. Actually, this equation is correct only for an infinite medium. As will be discussed later, the form for a finite medium changes the quantitative value of c but does not change the qualitative nature of the x-ray scattering effects to be considered. In any case, Huang took the displacements for the infinite medium as the basis of a model for a dilute solid solution in which the solute atoms have a different atomic radius than that of the solvent atoms. However, the model can be interpreted more generally if the solute atom is considered to be any defect of atomic dimensions which produces radial elastic displacements of the form $u=c/r^2$ and whose concentration is small.

Huang's analysis starts with the conventional expression for the scattering of x-rays by an assemblage of n atoms, namely,

$$Sf^2|\textstyle\sum_n \exp i(\boldsymbol{\xi}\cdot\mathbf{r}_n)|^2, \tag{1}$$

where S is the scattering calculated by Thomson for a single electron and f is the atomic scattering factor. The vector $\boldsymbol{\xi}$ is given by $2\pi/\lambda(\mathbf{n}-\mathbf{n}_0)$ where $\mathbf{n}_0$ and $\mathbf{n}$ are unit vectors in the direction of the incident and the scattered waves, respectively, and λ is the wavelength of the x-rays. The vector $\mathbf{r}_n$ locates the nth atom and is given by

$$\mathbf{r}_n=\mathbf{a}_n+\textstyle\sum_i \mathbf{u}(\mathbf{a}_n-\mathbf{a}_i), \tag{2}$$

where $\mathbf{a}_n$ locates the nth atom in the undistorted lattice and $\mathbf{u}(\mathbf{a}_n-\mathbf{a}_i)$ represents the vector displacement of the nth atom due to the ith defect. Substituting Eq. (2) in expression (1) and analyzing the problem on the basis of a statistically uniform distribution of defects, which distribution may fluctuate locally, Huang shows that the scattering becomes

$$Sf^2\textstyle\sum_n\sum_m \exp\{i\boldsymbol{\xi}\cdot[\mathbf{a}_n-\mathbf{a}_m+p\sum_i^{\text{all sites}}(\mathbf{u}(\mathbf{a}_n-\mathbf{a}_i)-\mathbf{u}(\mathbf{a}_m-\mathbf{a}_i))]\}\times\{1-\tfrac{1}{2}(1-\delta_{nm})(p-p^2)\times\textstyle\sum_i^{\text{all sites}}[\boldsymbol{\xi}\cdot(\mathbf{u}(\mathbf{a}_n-\mathbf{a}_i)-\mathbf{u}(\mathbf{a}_m-\mathbf{a}_i))]^2\}, \tag{3}$$

where p is the concentration of defects. The double summation arises in forming the square as required by expression (1). Substituting the explicit form for the displacements, namely, $\mathbf{u}=c\mathbf{r}/|\mathbf{r}|^3$, Huang then forms the indicated summations within the double summation signs in expression (3). These summations are resolved into three separate summations of which two are relatively simple but the third quite complex. Finally Huang obtains for the scattered intensity:

$$Sf^2\sum_n\sum_m\left[1-(1-\delta_{nm})(p-p^2)\frac{4\pi c^2|\boldsymbol{\xi}|^2}{3\rho_0 v}\right]\times\exp[i\boldsymbol{\xi}\cdot(\mathbf{a}_n-\mathbf{a}_m)]\left(1+\frac{4\pi cp}{3v}\right)+Sf^2(p-p^2)\times\frac{\pi c^2|\boldsymbol{\xi}|^2}{2v}\sum_n\sum_m(1-\delta_{nm})\left\{\frac{2\sin^2\Theta}{|\mathbf{a}_n-\mathbf{a}_m|}-\frac{(2\rho_0)^2\sin^2\Theta}{3|\mathbf{a}_n-\mathbf{a}_m|^3}-\frac{2(2\rho_0)^2\cos^2\Theta}{3|\mathbf{a}_n-\mathbf{a}_m|^3}\right\}\exp[i\boldsymbol{\xi}\cdot(\mathbf{a}_n-\mathbf{a}_m)]\left(1+\frac{4\pi cp}{3v}\right), \tag{4}$$

where v is the atomic volume, Θ the angle between $\boldsymbol{\xi}$ and $\mathbf{a}_n-\mathbf{a}_m$, and ρ_0 a cutoff which is difficult to evaluate exactly but which is of the order of atomic dimensions. In spite of the complexity of expression (4) certain features of the scattering are easy to recognize, remembering that the scattering by a stationary lattice without defects would be

$$Sf^2\sum_n\sum_m\exp[i\boldsymbol{\xi}\cdot(\mathbf{a}_n-\mathbf{a}_m)]. \tag{5}$$

The first term of expression (4) gives the usual sharp crystalline or Bragg reflections except that their positions and intensities are altered. The term $(4\pi cp)/(3v)$ in the exponent shows that the lattice has expanded isotropically as though a strain of this magnitude had been imposed upon the lattice. The quantity in the large square brackets can be regarded as the first two terms in the Taylor expansion of $\exp(-x^2)$. In this case, since $|\boldsymbol{\xi}|=(4\pi/\lambda)\sin\theta$, the quantity may be written in the form $\exp[-k(\sin^2\theta)/\lambda^2]$, where k is a constant, θ the Bragg angle, and λ the wavelength of the x-rays. Now the well known temperature factor for x-ray scattering by crystals, namely, $\exp[-B(\sin^2\theta)/\lambda^2]$, contains the same functions of θ and λ. However, the quantity B in the temperature factor depends on the temperature of the crystal whereas the constant k in the factor for localized static lattice defects is independent of temperature. This provides an important test for separating the x-ray effects due to the two factors. Owing to the functional similarity of the two factors, that for localized static lattice defects will be called the artificial temperature factor. Since the true temperature factor reduces the intensities of the Bragg reflections without altering their shapes, it is clear that the artificial temperature factor will produce the same effects.

The second term of expression (4) represents the diffuse scattering and is crudely analogous to temperature diffuse scattering, except that it is not, of course, temperature sensitive. The appearance of the same exponential factor as occurs in the first term assures that this scattering is located in the vicinity of the Bragg reflections, that is, surrounds the reciprocal lattice points. While the calculation of the detailed nature of

this diffuse scattering is complex, Huang shows that the diffuse scattering for dilute solid solutions is comparable in intensity and extension in reciprocal space to thermal diffuse scattering at low temperatures.

To summarize, the calculations of Huang predict theoretically the following x-ray scattering effects from a lattice containing a random distribution of defects which produce elastic displacements of the form $u=c/r^2$: (1) an isotropic expansion of the lattice; (2) an artificial temperature factor; (3) no broadening of the Bragg reflections; (4) diffuse scattering surrounding the reciprocal lattice points. It is the purpose of the next section to show that these effects are all observed experimentally.

Before proceeding to a discussion of the experimentally observed x-ray effects, however, it is appropriate to consider a matter which alters the results of Huang quantitatively in a way which may ultimately prove to be important, although it does not alter the results qualitatively. This matter concerns the use by Huang of the equation $u=c/r^2$ to express the displacements of the atoms around a defect. This equation is strictly applicable only to the case of an infinite medium. Eshelby[9] has given a very general analysis of the effect of neglecting the finite nature of the medium in problems using the elastic model for defects. In the problem solved by Huang it is stated by Eshelby that, due to the neglect of the surface term in the expression for the displacements, the constant c should be replaced by $c\gamma$, where $\gamma=3(1-\sigma)/(1+\sigma)$ and σ is Poisson's ratio. This increases the effective value of c by a factor of 1.8 or 1.5 depending on whether the value taken for Poisson's ratio is $\frac{1}{4}$ or $\frac{1}{3}$. Thus the magnitude of the effects predicted by Huang is increased by a significant amount.

EXPERIMENTAL

The isotropic lattice expansion and lack of line broadening predicted by Huang's calculations are confirmed by a very wide experience with solid solutions. Actually, as Tucker and Sampson[10] have pointed out, Huang's analysis, when corrected for the effect of the free surface term, predicts Vegard's law. This empirical law states that when two materials of the same crystal structure form a continuous series of solid solutions, the plot of lattice parameter *versus* composition is a straight line joining the parameters of the pure materials. While Vegard's law is surely an idealization, it forms a reasonable first approximation in many systems. The absence of line broadening in solid solutions is an experimental fact of long standing. Quite appreciable solubilities exist in some cases for which the radius mismatch between solvent and solute atoms amounts to 10 to 15%. Since the distortions around the solute atoms are certain to be large in such cases, the absence of line broadening provides a good test of Huang's calculations. No attempt has yet been made to look for the diffuse scattering, but the artificial temperature factor predicted by Huang's calculations for solid solutions has been reported by Coyle and Gale[11] in a gold-copper alloy.

While the lattice expansion, the absence of line broadening, and the presence of an artificial temperature factor in the gold-copper alloy studied by Gale provide evidence for the correctness of Huang's calcluations for the case of solid solutions, neutron irradiated single crystals provide the basis for a complete test of the calculations. In certain irradiated crystals all four of the predicted effects are found simultaneously. The reason that the effects are so pronounced in irradiated crystals is that the "strength" of the defects, as given by the constant c in $u=c/r^2$, is much greater for defects such as interstitials, vacancies, and foreign atoms (particularly the noble gases) than for substitutional foreign atoms in a solid solution. Since the constant c appears as a square in the terms involving the artificial temperature factor and the diffuse scattering in expression (4), the magnitude of these effects will depend strongly on the value of c.

Thus far the effects have been observed in irradiated single crystals of boron carbide, diamond, silicon carbide, and magnesium oxide. The case most studied has been that of boron carbide and a detailed account of these studies has been published elsewhere.[12] Most of the radiation damage produced in boron carbide is due to the reaction of the B^{-10} nucleus with thermal neutrons to form He^{-4} and Li^{-7} nuclei which then dissipate 2.3 Mev of kinetic energy in the lattice in ionization and bumping collisions. The lattice expansion and artificial temperature factor are highly anisotropic in irradiated boron carbide. However, it was possible to identify an anisotropic defect in the irradiated crystal structure by means of Fourier analysis of the Bragg intensities which accounted for the anisotropy of these effects. Work is in progress to generalize the calculations of Huang to the case of anisotropic defects. While this work is not yet complete, there is no indication that the effects will be qualitatively different from those predicted by Huang. The diffuse scattering effect is observed to a very extreme degree in irradiated boron carbide. Further, this diffuse scattering is temperature insensitive as the theory requires. The absence of broadening in the Bragg reflections is apparent in Laue photograms even for heavily irradiated crystals. Thus irradiated boron carbide exhibits all of the effects predicted by Huang.

The x-ray effects may also be seen in irradiated diamond, silicon carbide, and magnesium oxide crystals, although the artificial temperature factor is not very strong in magnesium oxide. Laue patterns illustrating the diffuse scattering produced in diamond, for example, are shown in Figs. 1 and 2. The diffuse spots

[9] J. D. Eshelby, J. Appl. Phys. **25**, 255 (1954).
[10] C. W. Tucker, Jr., and J. B. Sampson, Acta Metallurgica **2**, 433 (1954).
[11] R. A. Coyle and B. Gale, Acta Cryst. **8**, 105 (1955).
[12] C. W. Tucker, Jr., and P. Senio, Acta Cryst. **8**, 371 (1955).

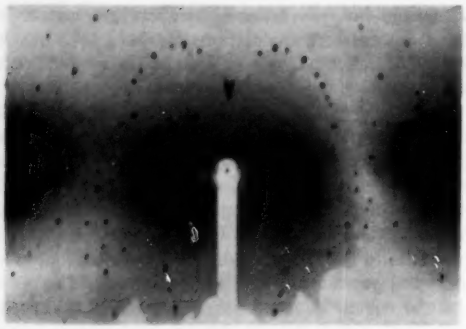


FIG. 1. Laue pattern of unirradiated diamond crystal using unfiltered copper x-radiation. Incident x-ray beam approximately parallel to threefold axis of crystal.

adjacent to and surrounding some of the Laue spots[13] in Fig. 2 show the diffuse scattering due to the defects produced by the irradiation. While the artificial temperature factor is not particularly apparent in Fig. 2, it is quite evident in Laue patterns (not shown) of irradiated silicon carbide. Lattice expansions[14] of the order of 1% are observed in irradiated diamond and silicon carbide, while that in magnesium oxide is about 0.1%. Thus all of the x-ray effects found in irradiated boron carbide are also observed in other irradiated crystal structures, showing that the effects are not peculiar to crystals undergoing fission type reactions.

The occurrence in boron carbide, diamond, silicon carbide, and magnesium oxide of the four x-ray effects predicted by the calculations of Huang provides good evidence that the elastic model used by Huang is a valid model for a crystal containing localized static lattice defects. It does not follow, however, that the effects will always be observed. For example, no diffuse scattering was found in irradiated silicon and germanium. But, since no lattice expansion has been observed in these substances,[14] it appears that there are very few defects of the type which produce the x-ray effects. Similarly, in metals the lattice expansions, in general, are quite small.[10] These observations fit in well with the widely accepted notion that the activation energy for the motion of isolated interstitials or vacancies in metals is rather low. Thus annealing of the main source of defects which could produce the x-ray effects occurs, in metals, at temperatures below room temperature. The crystals in which the x-ray effects have thus far been observed all possess high melting points which are indicative of tight bonding. It would not be expected that the activation energy for the decay of the defects in these crystals would be low. Actually, the anisotropic defect responsible for most of the x-ray effects in irradiated boron carbide was found to anneal in the temperature range of 700 to 900°C.[12]

Accepting a qualitative connection between the calculations of Huang and the observed x-ray effects in irradiated crystals, it becomes clear that it may be possible to learn something of the nature of the distortions in irradiated crystals from a detailed study of the x-ray effects. In this connection it appears most hopeful to study the diffuse scattering, for it is this scattering in the case of thermal effects which has been used many times[15] to determine the elastic constants of crystals and which recently has even been used to work back to the elastic vibration spectrum of copper.[16] However, it must be remembered that the diffuse scattering caused by static lattice defects is due to the distortion of the lattice around the defects rather than to the defects themselves. At the present time there is a discrepancy between theory and experiment regarding the diffuse scattering. Huang's theory predicts that the diffuse scattering around a reciprocal lattice point has roughly the shape of a lemniscate whose axis is parallel to the vector from the origin to the reciprocal lattice point and whose center lies at the reciprocal lattice point. Detailed work with boron carbide and other crystals has shown that the diffuse scattering has roughly the shape of an ellipsoid. This discrepancy may mean that the displacements around the defects deviate from the form $u=c/r^2$ assumed by Huang or that replacement of c by $c\gamma$ as suggested by Eshelby[9] is not valid for the diffuse scattering term. Further theoretical and experimental work is necessary to work out the quantitative aspects of the diffuse scattering.

CONCLUSIONS

The theoretical calculations of Huang show that a lattice containing a random distribution of elastic

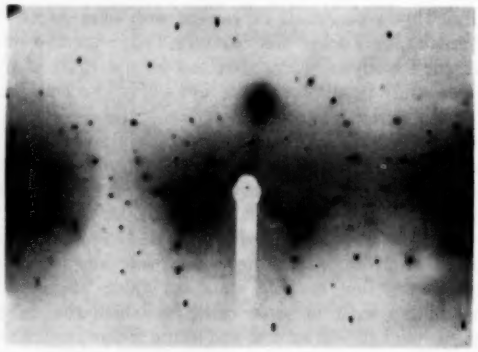


FIG. 2. Laue pattern of neutron irradiated diamond crystal using unfiltered copper x-radiation. Integrated thermal neutron flux 2×10^{20} neutrons/cm^2. Incident x-ray beam approximately parallel to threefold axis of crystal.

[13] Only those planes at or near the Bragg setting for reflection of the very strong characteristic $K\alpha$ or $K\beta$ radiation from the target show the diffuse scattering.

[14] W. Primak, Phys. Rev. 95, 837 (1954).

[15] See, for example, G. N. Ramachandran and W. A. Wooster, Acta Cryst. 4, 431 (1951).

[16] E. H. Jacobsen, Phys. Rev. 97, 654 (1955).

singularities which produce displacements of the form $u=c/r^2$ will scatter x-rays with the following effects: (1) an isotropic expansion of the lattice; (2) an artificial temperature factor; (3) no broadening of the Bragg reflections; (4) diffuse scattering surrounding the reciprocal lattice points.

While the lattice expansion and absence of line broadening have been widely observed in solid solutions, the artificial temperature factor has been observed only once and the diffuse scattering has not yet been observed in solid solutions. However, all four of the effects have been observed in certain neutron irradiated crystals, namely, boron carbide, diamond, silicon carbide, and magnesium oxide. The presence of the effects in these crystals is attributed to the defects produced during irradiation. The fact that the effects are readily observed in irradiated crystals is due to the much greater "strength" of the defects produced by irradiation, *viz.*, interstitials, vacancies, and foreign atoms.

The qualitative agreement between the calculations of Huang and the observed x-ray effects leads to the conclusion that the representation of localized static defects as elastic singularities producing displacements of the form $u=c/r^2$ is a valid first-approximation model. However, it is suggested that it may be possible to work back from the observed diffuse scattering in irradiated crystals to obtain more precise information regarding the distortions around the defects.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. J. B. Sampson and Dr. W. M. Cashin for their interest in this work.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Luminescent Centers in ZnS:Cu:Cl Phosphors

R. Bowers and N. T. Melamed
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania

(Received April 26, 1955)

Measurements of the magnetic susceptibility and emission spectra have been made on zinc sulfide activated by copper and/or chlorine. The results of these measurements demonstrate that the luminescent centers are diamagnetic in the absence of optical excitation, and that of the five previously reported emissions, only three are distinct. These are: a blue emission resulting from the addition of chloride, a green emission when copper and chloride are simultaneously present, and a red emission when copper alone is present. The results are discussed on an ionic model, and on a semiconductor model. We conclude that the blue emission is due to zinc vacancies, the green emission to substitutional copper, and the red emission to interstitial copper.

INTRODUCTION

IN this paper we shall attempt to correlate the reported emissions in ZnS activated by copper with particular states of the activator, especially the valence of the activator and its position in the lattice.

The absorption of a quantum of ultraviolet radiation by the host crystal of ZnS is usually assumed to result in the formation of an electron and a hole. These may annihilate one another by giving up their energy to the lattice as phonons, or they may recombine at an impurity center with a characteristic emission. The electronic transitions are governed by interactions between the impurity and the lattice; it is to be expected, therefore, that the resulting emission would depend on the state of ionization of the center. Furthermore, the energy levels may well depend on the crystallographic phase of the host crystal, and on whether the impurity is incorporated interstitially or substitutionally.

Information regarding the nature of impurity centers in impurity activated phosphors has been obtained for several manganese-activated inorganic phosphors by means of static susceptibility measurements,[1] and for a variety of phosphors by means of paramagnetic resonance.[2] In the case of ZnS:Cu, however, paramagnetic resonance measurements failed to yield any information because of the absence of an observable resonance absorption. The difficulties encountered in the resonance measurements, in being unable to distinguish between diamagnetism and a paramagnetism which results in a very broad absorption spectrum, are absent in the measurements of static susceptibility.

In the experiments to be described here, we have measured the emission spectra of a number of Cu-activated ZnS phosphors, prepared under various conditions, while at the same time, the phase has been determined from x-ray measurements. The state of ionization of the unexcited phosphors has been determined from measurements of the magnetic susceptibility as a function of temperature. Zinc sulfide is diamagnetic and, if the copper were introduced as Cu^+, it would remain purely diamagnetic and its susceptibility would be substantially independent of temperature. On the other hand, the presence of Cu^{++} or Cu^0

[1] S. Larach and J. Turkevich, Phys. Rev. **89**, 1060 (1953). P. D. Johnson and F. E. Williams, J. Chem. Phys. **17**, 435 (1949).

[2] W. D. Hershberger and H. N. Leifer, Phys. Rev. **88**, 714 (1952).

would result in a temperature-dependent paramagnetic component in the susceptibility.

Five principal emissions of the ZnS:Cu:Cl system have been reported. The chemical and optical properties of this system have been the subject of much investigation.[3] It is well established that the presence of chloride during the preparation of zinc sulfide results in a blue emission band under near-ultraviolet excitation, having its peak at approximately 4400 A for the hexagonal phase and at 4600 A for the cubic form.[4] The addition of small quantities of copper together with chloride produces a strong green emission with peaks at 5200 A and 5400 A for the hexagonal and cubic forms respectively.[5] At low copper concentrations,[6] the blue and green peaks are both present, the higher concentrations of copper favoring the green emission, while at the same time diminishing the intensity of the blue. The green emission reaches its maximum intensity at 0.009 to 0.01 mole percent copper and at this concentration, the blue emission is virtually absent under ordinary conditions of excitation.

With the addition of copper beyond 0.01%, the green emission diminishes in intensity and a blue emission reappears. This blue emission reaches its maximum intensity at 0.06% copper and diminishes at higher concentrations.[6a,7] At a concentration of 0.3% copper, Bube[5] has reported a new green emission, having a maximum at 5200 A, but of a considerably broader band width. The foregoing emissions are all presumed to occur in the presence of an excess of chloride. Recently, Froehlich[8,9] has described a broad red emission produced when copper is added to zinc sulfide and prepared in the absence of any additional impurity such as chloride. The peak wavelengths of all of the emissions and their optimum copper concentrations are summarized in Table I.

TABLE I. Impurity concentration and representative centers proposed in the literature for ZnS.

Impurity	Emission color	Wavelength at maximum A	Proposed centers
Cl	blue I	4400	interstitial Zn^+, Zn vacancies
Cu(0.01%):Cl	green I	5200	Cu^+ interstitial, or substitutional
Cu(0.06%):Cl	blue II	4400	$Cu_2{}^+Cl$
Cu(0.3%):Cl	green II	5200	subt. Cu^+
Cu	red	6700	? Cu^{++}

[3] See for example H. W. Leverenz, *An Introduction to the Luminescence of Solids* (John Wiley and Sons, Inc., New York, 1950), Chap. 5.

[4] F. A. Kroger and J. E. Hellingman, J. Electrochem. Soc. **93**, 156 (1948); also S. Rothschild, Trans. Faraday Soc. **42**, 635 (1946).

[5] R. H. Bube, Phys. Rev. **80**, 655 (1950), see also Leverenz, reference 3, pp. 208, 209.

[6] All concentrations are designated in molar percent.

[6a] Kroger, Hellingman, and Smit, Physica **15**, 990 (1949).

[7] F. A. Kroger and N. W. Smit, Physica **16**, 317 (1950).

[8] H. Froelich, J. Opt. Soc. Am. **42**, 982, (1952).

[9] H. Froelich, J. Electrochem. Soc. **100**, 496 (1953).

It has been stated that the green emission at low copper concentrations results from Cu^+, and that the role of the chloride is to provide charge compensation in a lattice of $Zn^{++}S^{=}$.[10] The blue emission occurring when chloride alone is present, had formerly been attributed to Zn^+ centers[11] but recently has been attributed to Zn^{++} vacancies with an electron missing from one of the adjacent $S^=$ ions.[12] The presence of Zn^+ should be detectable in our experiment by its paramagnetic susceptibility.

The origin of the blue emission appearing at 0.06 mole percent copper is not well established. It has been suggested that this blue emission is due to a complex of the form $[Cu_2{}^+:Cl^-]$.[6,13] These centers should be paramagnetic owing to the presence of an unpaired electron spin. A similar lack of knowledge exists with regard to the red emission which appears when no chloride is present. Because of the absence of a charge compensating ion such as Cl^-, one might be led to expect Cu^{++} as the responsible center; hence these specimens may exhibit paramagnetic behavior.

EXPERIMENTAL PROCEDURES AND RESULTS

Preparation of the Specimens

Specimens were prepared by heating together pure precipitated zinc sulfide (R.C.A. LM476) to which requisite quantities of activator had been added. The copper activator was added as a solution of $Cu(NO_3)_2$. The components were mixed, dried, then thoroughly ground to insure proper mixing. Chloride was added as ammonium chloride, in large excess, usually 3%. The precise amount of chloride which remains in the lattice after firing is not certain; it is for this reason that the chloride concentrations are not included in the figures. Bube[14] and Kroger[15] have reported analyses for the chloride remaining; in general the chloride concentration increases with the amount of copper added. Without copper, however, a quantity of chloride is incorporated and produces the aforementioned blue emission. There is more certainty regarding the copper concentrations when chloride is present. Chemical analyses, and the continuous changes in physical properties with the copper concentration lead us to believe that, at least to 0.1% copper, all, or a substantial portion of the added copper is incorporated.

All chloride-free specimens were heated in an atmosphere of hydrogen sulfide and precautions were taken to exclude air. Chloride-containing specimens were fired in an atmosphere consisting of a mixture of hydrogen sulfide and ammonium chloride. The firing temperature and time depended on whether the hex-

[10] F. A. Kroger and J. Dikhoff, Physica **16**, 297 (1950).

[11] J. T. Randall, Trans. Faraday Soc. 8 (1939).

[12] F. A. Kroger and H. J. Vink, J. Chem. Phys. **22**, 250 (1954).

[13] F. A. Kroger, Brit. J. Appl. Phys., Supplement No. 4, 58–62 (1955).

[14] R. H. Bube, J. Chem. Phys. **19**, 985 (1951).

[15] F. A. Kroger and J. E. Hellingman, J. Electrochem. Soc. **95**, 68 (1949).

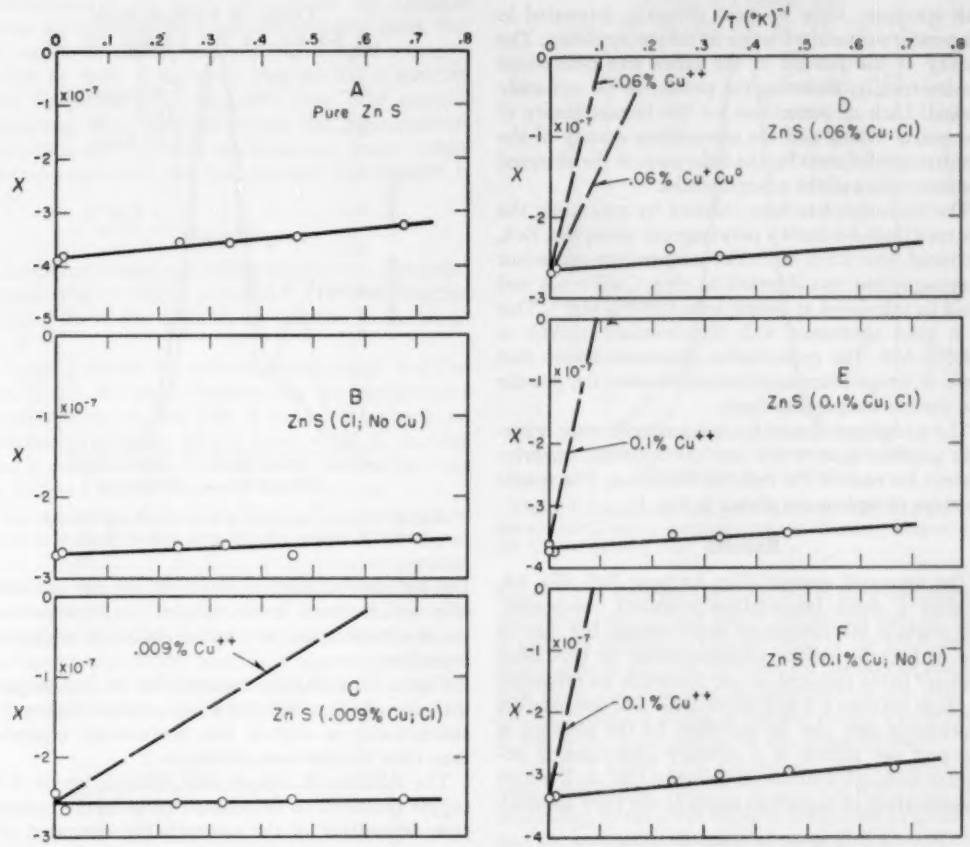


FIG. 1. Temperature dependence of the susceptibility of ZnS with various concentrations of copper and chloride. Emission bands corresponding to the compositions shown are: *B*, "blue I"; *C*, "green I"; *D*, "blue II"; *F*, red.

agonal or cubic phase was desired. For the hexagonal specimens reported on here, the firing was for 30 minutes at 1150°C. The red-emitting specimens were fired for considerably longer periods at the above temperature as well as at higher temperatures, in an attempt to produce the hexagonal form, but the products were mixtures of cubic and hexagonal phases.

Measurement of the Magnetic Susceptibility

The susceptibility of the powder specimens has been measured from 300°K down to 1.5°K using the Gouy method. A Gulbransen type microbalance capable of measuring a force of 1 μg with a 3-g load has been used to measure the force on the specimen.[16,17] Because of the sensitivity of the balance and the wide temperature range covered, it is possible to separate the temperature-dependent paramagnetism which would correspond to 0.0001% molar Cu^{++} in ZnS.

[16] E. A. Gulbransen, Rev. Sci. Instr. **15**, 201 (1944).
[17] R. Bowers and E. A. Long, Rev. Sci. Instr. **26**, 337 (1955).

The phosphors to be measured were placed in thin-walled quartz tubes; they were evacuated to 10^{-6} mm and heated to 300°C for one hour in the vacuum in order to remove paramagnetic oxygen. The tubes were then filled with just less than one atmosphere of pure helium before sealing to ensure thermal equilibrium during the measurements.

The actual susceptibility measured is that of the phosphor and quartz container. The susceptibility of the quartz glass has been measured from 300°K to 1.5°K and has been found to be suitable for our work because its variation with temperature is very small.

The correction due to the quartz container was often as high as 30%, as calculated from a measured average cross-sectional area of the container. However, the quartz tubing was not uniform and the correction itself could not be calculated to better than about 30%. We did not consider it worthwhile to devote much attention to an accurate evaluation of this almost temperature-independent correction in the case of

each specimen, since we were primarily interested in temperature-dependent terms in the susceptibility. The density of the powder in the tubes was determined gravimetrically, assuming the powder to be uniformly packed. Lack of corrections for the nonuniformity of the quartz tubing and the nonuniform density of the powders may account for the differences in the observed absolute values of the susceptibilities.

The technique has been checked by measuring the susceptibility of a known paramagnetic phosphor, ZnS, activated with 0.1% Mn^{2}. The temperature-dependent paramagnetism was observed to obey Curie's law and could be calculated as arising from 0.093% Mn^{++}. This is in good agreement with the chemical analysis of 0.095% Mn. The quantitative agreement shows that there is temperature equilibrium between the powder and the low temperature bath.

The phosphors chosen for measurement were, wherever possible, ones which had the optimum impurity content for each of the reported emissions. The results for these phosphors are shown in Fig. 1.

Results

The measured susceptibility for pure ZnS, Fig. 1A, exhibits a small temperature-dependent component. The origin of this component is not certain, but may be due in part to surface paramagnetism or crystalline defects[18] in the particles, whose diameters are estimated to range between 0.1 and 10 microns. The temperature dependence may also be explained by the presence of one part per million of a strongly paramagnetic impurity, although there is no evidence that such a high concentration of impurities exists in the pure material.

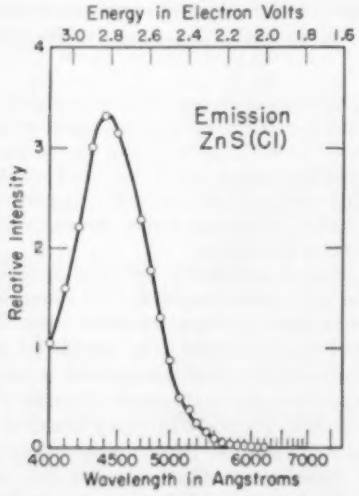


FIG. 2. Emission spectrum of ZnS containing chloride but no copper; designated in text as "blue I".

[18] P. W. Selwood, *Magnetochemistry* (Interscience Publishers, Inc., New York, 1953), p. 253.

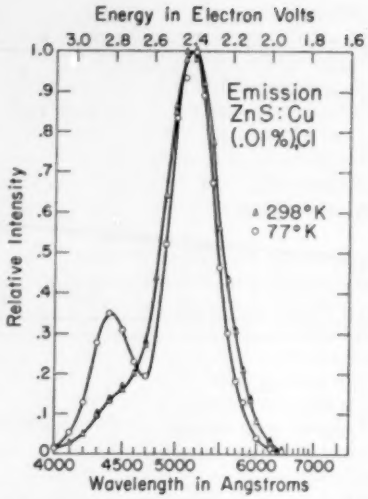


FIG. 3. Emission spectrum of ZnS containing chloride and 0.01% copper; referred to in text as "green I".

For our present purpose, however, we are concerned only with increases in the temperature dependence of the susceptibility as the result of deliberate addition of impurities.

Figure 1B shows the susceptibility of ZnS prepared with the addition of chloride but without copper. The susceptibility is slightly less temperature dependent than even the pure zinc sulfide.

The addition of copper with chloride up to 0.1% copper produced no detectible change in the temperature dependence of the susceptibility compared with the pure zinc sulfide. This is illustrated in Figs. 1C, D, and E for samples with increasing copper concentration. In each of these figures, the dotted line represents the expected slope of the added copper had it been present as Cu^{++} or $Cu^{0}Cu^{+}$.

Figure 1F shows the temperature dependence of the susceptibility for a specimen containing 0.1% copper prepared without chloride, and exhibiting a red emission. As in the previous specimens, no change in the original susceptibility is observed.

We conclude, therefore, that neither the addition of copper or chloride, nor the addition of copper together with chloride in concentrations necessary to produce the reported emissions, results in any observable increase in the temperature dependence of the susceptibility of zinc sulfide. Hence, the impurities are present in forms which are diamagnetic.

Emission Spectra

Emission spectra were obtained by means of a Beckman spectrophotometer with a 931A photomultiplier replacing the Beckman light source. The calibration included the sensitivity of the phototube and the

transmission of the spectrophotometer. All emission curves are reported on a relative-intensity basis. For the emission spectra, the phosphors were normally excited by 3650 A radiation from an A-H4 mercury lamp containing a Corning 5860 filter. The spectrophotometer band width at which the measurements were taken was kept at a minimum and, except where otherwise indicated, the half-intensity band width is 40 A.

Results

Pure, unactivated zinc sulfide exhibits an ultraviolet emission but no visible emission.[19] The blue emission resulting from the addition of chloride is shown in Fig. 2.

Figure 3 shows the emission spectrum of ZnS:Cu-(0.01%):Cl. At room temperature the predominant emission is due to the 5200 A band, with almost no contribution from the 4400 A band. At 77°K the blue band is considerably enhanced, as can be expected from the Schoen-Klasens[20] theory of hole migration.

The emission of ZnS:Cu(0.065):Cl, the optimum concentration for the blue II emission, is shown in Fig. 4. A difference in the shape of the emission spectrum of the blue band in the copper-containing specimen and the blue band of ZnS:Cl is seen in the more rapid decrease in the intensity of the former on the short-wavelength side of the band. The approximate shape of the absorption band introduced by copper in the tail of the lattice absorption band has been obtained by

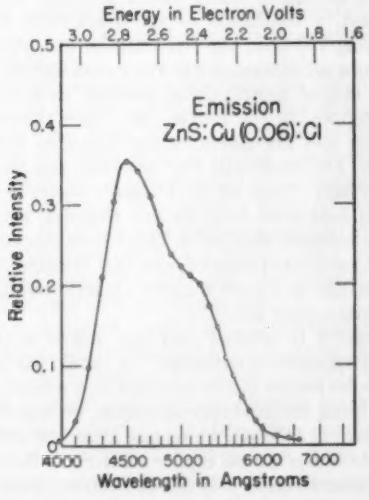


FIG. 4. Emission spectrum of ZnS containing chloride and 0.06% copper, referred to in text as "blue II". A small green I component of emission is present.

[19] F. A. Kroger, Physica 7, 1 (1940).
[20] M. E. Wise and H. A. Klasens, J. Opt. Soc. Am. 38, 226 (1948); Klasens, Ramsden, and Quantie, J. Opt. Soc. Am. 38, 60 (1948); H. A. Klasens, Nature 158, 306 (1946); M. Schoen, Z. Physik 119, 463 (1942).

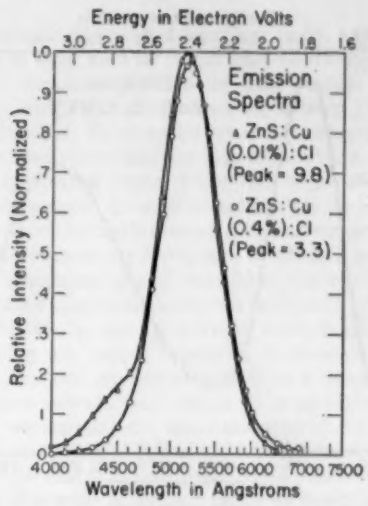


FIG. 5. Comparison of the green emission bands obtained with low and with high copper concentrations. These are designated in the text as "green I" and "green II".

Gisolf.[21] The difference in the shape of the blue emission spectra of ZnS:Cl and ZnS:Cu(0.06):Cl can, at least in part, be attributed to absorption of the short-wavelength side of the blue emission by the copper absorption band. An apparent displacement of the peak of the blue emission of Fig. 4 towards longer wavelengths is due to the superposition of the green emission.[22]

Figure 5 shows on a normalized ordinate scale the room-temperature emission of ZnS:Cu(0.4%):Cl, corresponding to the band we have designated as green II, and for comparison the emission of ZnS:Cl(0.01):Cl. Except for the presence of a small amount of blue I in the 0.01% copper specimen no differences are observed which can be attributed to a broadening of the emission at high copper concentrations.

Figure 6 shows the room-temperature emission of ZnS:Cu(0.01) prepared without chloride. The broad red emission has its maximum at 6700 A. Also present is a blue emission similar to the blue emission present in zinc sulfide prepared with chloride, but without copper. The presence of the blue emission together with the red emission has been verified for all of our specimens ranging in copper concentration from 0.0001% Cu to 1% Cu. In some specimens, only the red emission appears at room temperature; the presence of blue centers can readily be established, however, by cooling the specimen to 77°K.

DISCUSSION

There has been considerable speculation regarding the number of different emission centers and their

[21] Gisolf, De Groot, and Kroger, Physica 8, 805 (1941).
[22] J. M. Vanderbilt and C. Henrich, Appl. Spectroscopy 7, 171 (1953).

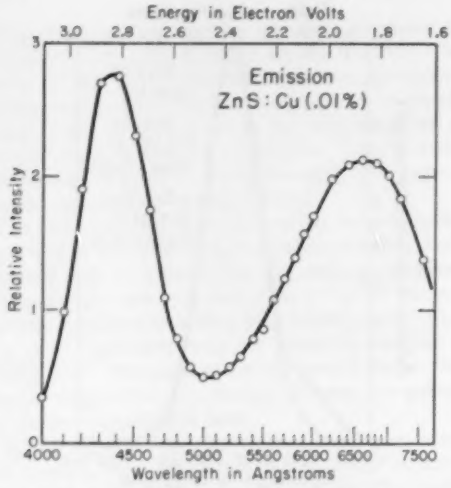


FIG. 6. Emission spectrum of ZnS containing 0.01% copper but no chloride, showing the long-wavelength red emission with its associated blue emission. The intensity of the blue emission in this specimen is greater than in many similar specimens.

constitution, in zinc sulfide activated by copper and by chloride. Our results indicate that only three emissions are to be distinguished out of the five reported: a blue emission produced by chloride alone, a green emission due to the simultaneous addition of copper and chloride, and a red emission due to the presence of copper without chloride. Emission spectra obtained for the blue-emission specimens with high copper concentration (0.06 molar percent) were similar to those obtained with specimens prepared with only chloride. At very high (0.34%) copper concentrations, no broad green emission was observed. The green emission retained its shape to the highest concentrations to which we went; the only noticeable changes were in relative intensity. This latter observation confirms Kroger's[23] view that no new emission appears at the higher copper concentrations.

The susceptibility measurements obtained on all of the specimens described above indicate that the concentration of paramagnetic centers introduced by copper or chloride was, in each case, less than 0.001 molar percent. Thus, it is not likely that, in the ground state, any configuration of copper other than Cu^+ is present, as opposed to centers of the type Cu_2^+, Cu^{++}, or Cu^0. If any of the latter are present, their concentrations are too low to be correlated with the observed emissions.

On the basis of the above results, several conclusions can be drawn as to the probable nature of the luminescent centers. We shall discuss these results first assuming an ionic model, and then indicate how they may be interpreted using an energy band model.

[23] F. A. Kroger, J. Chem. Phys. **20**, 345 (1952).

Regarding the blue emission, the two most obvious possibilities are that this emission occurs at a Cl^- site, or in the region of a zinc vacancy. (The possibility of Zn^+ being responsible for the blue emission is ruled out because such a center would be paramagnetic.) Since the same blue emission can be produced by a variety of impurities other than Cl^-, such as Br^-, I^-, Al^{+++}, Sc^{+++},[6] all of which would produce Zn vacancies, as well as by heating the pure material to a sufficiently high temperature,[5] the evidence favors Kroger's recent contention[12] that Zn vacancies are responsible for this emission. It is, however, in disagreement with Krogers picture of a blue center as that of a zinc vacancy with one electron missing from an adjacent sulfur ion, inasmuch as the missing electron would produce a paramagnetic center.

The situation regarding the green and red emissions is somewhat less apparent. Because of the difference in charge between Zn^{++} and Cu^+, the incorporation of Cu^+ substitutionally in a lattice such as ZnS requires the simultaneous incorporation of a monovalent negative ion, such as Cl^-, Br^-, or I^-, or a tripositive ion, such as Al^{+++}, or sulfur vacancies. Since one of these impurities must be present with copper in order to produce the green emission, we conclude that in this case the copper is present substitutionally. Furthermore, since the emission is the same for any one of the above impurities, direct association between them and the copper can be ruled out, and the green emission must arise from the copper in a substitutional site having a normal or nearly normal zinc and sulfur environment, a conclusion in agreement with Kroger. This being the case, the red emission can arise from copper in a substitutional site whose environment differs from a normal substitutional position, or from interstitial copper. In the former case, since charge compensating ions are absent, sulfur vacancies would be required. For interstitial Cu^+ however, zinc vacancies are necessary. Since we have always observed a blue emission associated with the red emission, similar to the blue emission obtained in ZnS:Cl, we conclude that zinc vacancies are produced, and that therefore the red emission due to copper without chloride results from interstitial copper ions.

Because it is believed that zinc sulfide is at least partially non-ionic in character,[24] we should like to show that similar results can be obtained on a semiconductor model. Using the band approximation, we may describe the system as follows: We replace a neutral zinc atom in sulfide with a neutral copper atom. Since the neutral copper atom contains only one 4*s* electron, there will be a deficiency of one electron per copper atom, and positive holes will appear in the valence band. In semiconductor terminology, copper will act as an acceptor impurity. At high temperatures, with the acceptors ionized, the copper atoms will each have one additional

[24] Y. K. Syrkin and M. E. Dyatkina, *Structure of Molecules* (Butterworths Scientific Publications, London, 1950), p. 340.

electron, and will thus be diamagnetic; the system as a whole will be paramagnetic, however, due to the presence of free holes. At sufficiently low temperatures, the holes will be associated with the copper atoms, and the system will remain paramagnetic. Similarly, we assume that the replacement of sulfur by chlorine will produce donor centers. In this case also the system will be paramagnetic, whether or not the centers are ionized.

If we now introduce equal numbers of donor and acceptor centers by adding equal amounts of copper and chlorine, the ZnS will remain intrinsic; the donor electrons produced by chlorine will have annihilated the holes produced by copper, and the system will be diamagnetic at low temperatures. It is thus noted that in this situation, copper will have one more electron, and chlorine one less electron than in the neutral atomic configurations.

It can be shown in a quantitative fashion[25] that in many semiconductors the number of acceptor levels will always very nearly equal the number of donor levels. The condition for this is that the energy gap be large with respect to the energy necessary to produce a defect such as a vacancy. If one adds only a donor impurity to such a semiconductor, for example, it is energetically economical for the system to produce an equivalent number of *physical* acceptor centers by means of the appropriate kind of vacancies. This is the situation presumed to exist in the cases discussed here. It is to be expected, therefore, that if copper and chlorine are both present in the preparation, they will both be assimilated, in any relative proportion which will produce nearly equal numbers of donor and acceptor states.

The above description, involving the simultaneous presence of copper and chlorine, is the situation we believe to exist when the green emission is observed. The need for chlorine or a similar donor impurity, together with the copper, to produce this emission leads us to conclude that the copper is substitutional. The absence of any appreciable difference between the green emission produced with any other donor such as bromine or iodine indicates that the emission is associated mainly with the copper impurity.

As was indicated earlier, if substitutional chlorine donor atoms are introduced by themselves they will result in the production of acceptor levels; in the absence of a second impurity, the acceptor levels can be provided by vacancies. That the blue emission is a property of the acceptor states and not of the chlorine impurity is indicated by the fact that the same blue emission occurs with either bromine or iodine.[10]

Although direct evidence is lacking concerning the defect responsible for these acceptor levels, arguments similar to those used for copper lead us to assign them to zinc vacancies. This choice finds support in (a) analogy with other semiconductors, notably PbS and PbSe, in which Pb vacancies behave as acceptors, and (b) the observation that the blue emission can be produced by heating impurity-free ZnS to a sufficiently high temperature. In order to account for the diamagnetism when chlorine is present, each zinc vacancy must produce two acceptor levels each of which is occupied in the luminescent ground state by an electron.

The same blue emission which is produced by chlorine is observed in the specimens having a red emission, and containing only copper impurities. In these red specimens, therefore, the copper behaves as a donor. Since we believe substitutional copper to be an electron acceptor, we attribute the donor character of the copper in these specimens to interstitial copper atoms.

As in the case of substitutional copper, the interstitial copper, if introduced without vacancies would be paramagnetic due to the one 4*s* electron. Because of the simultaneous presence of acceptor levels in the form of zinc vacancies, the copper atoms are ionized, and the system remains diamagnetic.

The semiconductor description leads us to certain conclusions which were similar to those obtained on an ionic model: the green emission is attributed to substitutional copper, the red emission to interstitial copper, and the blue emission to zinc vacancies. There are differences, however, in the detailed description of these centers. On the ionic model we would describe both the interstitial and substitutional copper as Cu^{+}. This implies that in their optically unexcited states interstitial and substitutional copper impurities are alike in that they each have one electron less than the neutral atom. On the semiconductor model, the two are quite different; the substitutional copper has gained an electron while the interstitial has lost one. Other differences also result from the different assumptions made in the two descriptions. Charge compensation in the ionic model is replaced by an equality of donors and acceptors. By means of the latter principle, we would not expect to find copper present in ZnS as Cu^{++}, whereas Cu^{++} should be possible on the basis of charge compensation alone, and should require no coactivation nor vacancy.

ACKNOWLEDGMENTS

The authors are grateful to Dr. E. N. Adams and Dr. R. L. Longini of the Westinghouse Research Laboratories, and to Professor A. M. Portis of the University of Pittsburgh for many valuable discussions and criticisms. We are also indebted to Miss Rita Reba for her technical assistance.

[25] R. L. Longini, Bull. Am. Phys. Soc. **30**, No. 3, 36 (1955).

Anisotropy and Magnetostriction of Some Ferrites

R. M. Bozorth, Elizabeth F. Tilden, and Albert J. Williams
Bell Telephone Laboratories, Murray Hill, New Jersey

(Received May 12, 1955)

Magnetic crystal anisotropy and magnetostriction have been measured in various single crystals of ferrites having compositions represented approximately by MFe_2O_4, where M stands for Mn, Fe, Co, Ni, and Zn in various proportions. Special attention is given to heat-treatment in a magnetic field.

The magnetic anisotropy of cobalt ferrite at room temperature is as high as 4×10^6 ergs/cm³. Magnetostriction is as high as 800×10^{-6}. Magnetic anneal is effective at temperatures as low as 150°C, and causes the hysteresis loop to become square. In polycrystalline material the response to magnetic anneal is a maximum at compositions intermediate between $CoFe_2O_4$ and Fe_3O_4.

The constants for the various specimens are tabulated. Values of the anisotropy constants of $MnFe_2O_4$ at 20°C and −196°C are the same as those determined from ferromagnetic resonance experiments. At −196°C the constant for $Ni_{0.75}Fe_{2.25}O_4$ differs markedly from that determined by ferromagnetic resonance; this is to be expected from the relaxation phenomena observed by Galt, Yager, and Merritt.

IT is well known that some magnetic materials respond to heat treatment in a magnetic field: after cooling from an elevated temperature in the presence of a magnetic field, the domains are found to have a preferred direction parallel to the field, and the magnetic properties are correspondingly anisotropic. The chemically combined mixed oxides of cobalt and iron (cobalt ferrites) have been known to be responsive to such a heat treatment since the work of Kato and Takei.[1]

In an attempt to discover the explanation of this phenomenon, and to study the behavior of ferrites in other ways, experiments have been made on the crystal anisotropy and magnetostriction of several cobalt ferrites in both single-crystal and polycrystal forms. Special attention is given to the effect of heat-treatment. Measurements of these properties have also been made on a variety of other ferrites.

SPECIMENS

Most of the single-crystal specimens, grown by the flame fusion method, were obtained from Dr. G. W. Clark of Linde Air Products Company. Others were grown by J. P. Remeika of the Bell Laboratories; this work will be published shortly. The former crystals contained an excess of Fe_3O_4 over the formula $CoFe_2O_4$, the latter were nearly stoichiometric.

Most of the experiments on the effect of magnetic anneal on the crystal anisotropy were carried out on crystals of cobalt zinc ferrite, $Co_{0.22}Zn_{0.24}Fe_{2.18}O_4$. This formula and others given below are based on chemical analysis of the crystals measured for metallic elements, the remainder being assumed to be oxygen. If we assume the ions to be Zn^{++}, Co^{+++}, and Fe^{+++}, this can be written $(ZnO)_{0.22}\cdot(Co_2O_3)_{0.14}\cdot(Fe_2O_3)$; valences cannot be satisfied with Co^{++} instead of Co^{+++}. Single crystals of $Co_{0.73}Fe_{2.27}O_4$ were also examined; these correspond to $(CoO)_{0.73}\cdot(FeO)_{0.27}\cdot(Fe_2O_3)_{0.95}$. Continued heating in air at temperatures up to 600°C did not change the weights of the specimens.

Most of the polycrystalline specimens were prepared under the direction of J. W. Nielsen and were fired at 1200°C to 1250°C in an atmosphere of O_2 or CO_2. They were intended to cover the range of solid solutions extending from $CoFe_2O_4$ to Fe_3O_4.

The composition of these and other materials are given in Tables I and II.

CRYSTAL ANISOTROPY

Measurements of the torque in a saturating magnetic field were made on single crystal disks weighing less than 0.1 g, by using well-known techniques.[2] Figure 1 shows torque curves for a single crystal disk of cobalt zinc ferrite cut parallel to a (100) crystal plane, θ being the angle between the saturating field (or saturation magnetization, M_s) and the [001] crystal direction. When the specimen is heated at 150°C for three days with no magnetic field present ($H_T=0$) we obtain curve A, which shows the expected symmetry for a cubic crystal,

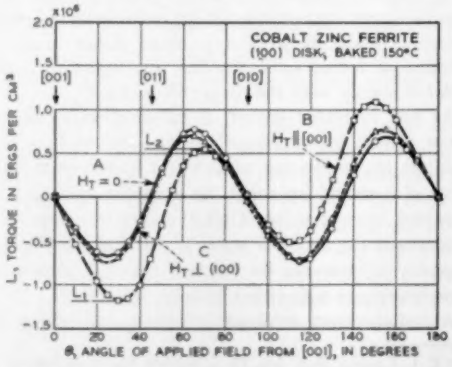


FIG. 1. Torque curves, showing the effect of magnetic anneal parallel and perpendicular to the (100) plane of measurement.

[1] Y. Kato and T. Takei, J. Inst. Elec. Engrs. (Japan) 53, 408 (1933).

[2] R. M. Bozorth, *Ferromagnetism* (D. Van Nostrand Company, Inc., New York, 1951), p. 556.

with easy directions parallel to the cubic axes. The torque is given closely by the theoretical relation $L=-\frac{1}{4}K_1 \sin 4\theta$, with the crystal anisotropy constant equal to $K_1=+1.5\times10^6$ ergs/cm³.

After maintaining the specimen for the same temperature and time in the presence of strong magnetic field ($H_T=9000$ oersteds) parallel to [001], curve B is obtained. This shows that the stability of the easy direction parallel to H_T has been enhanced, while that of the easy direction at right angles to H_T, [010], has been diminished. The difference in the crystal anisotropy energy between the [001] and [010] directions is now 0.5×10^6 ergs/cm³, whereas it was formerly zero.

When the specimen is annealed with $H_T\perp$ (100) plane, it is noted that the torque measured in this plane is practically the same (curve C) as when no field was present during annealing (curve A).

Measurements of the torque on a disk cut parallel to a (110) plane of the same crystal are shown in Fig. 2 for

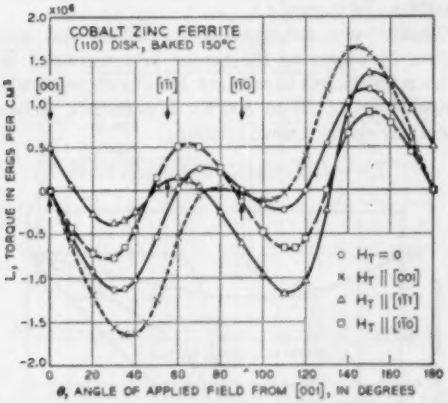


FIG. 2. Effect of magnetic anneal in various directions in the (110) plane.

various directions of H_T. The effect of the direction of the field present during heat treatment is clearly shown. Here we find not only differences in the relative energies, depending on the direction of H_T, but also a shift in the directions of easy magnetization. These occur at $L=0$, when $dL/d\theta<0$, and are parallel to the [001] and [1̄10] directions when $H_T=0$.

The results given above can be described by the superposition of an uniaxial anisotropy on the normal cubic anisotropy. The whole crystal anisotropy can then be expressed as follows:

$$E=K_1(\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2)+K_T\sin^2(\theta-\theta_T), \quad (1)$$

the α's being the direction cosines of the saturation magnetization, M_s, with respect to the crystal axes, θ_T the angle between H_T and the nearest crystal axis, and θ the angle between M_s and [001]. The torque is then

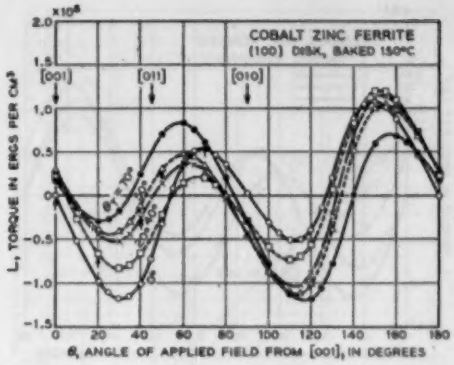


FIG. 3. Effect of magnetic anneal in various directions in the (100) plane, θ_T being angle between [001] direction and direction of field during anneal.

$L=-dE/d\theta$, and for the (100) plane is

$$L=-\tfrac{1}{4}K_1\sin 4\theta-K_T\sin 2(\theta-\theta_T), \quad (2)$$

when H_T lies in the plane.

To test this relation further, measurements were made with the (100) disk with H_T varied in approximately 10° steps from [001] to [010] in (100). The resulting torque curves (Fig. 3) were analyzed by evaluating the components containing $\sin 2\theta$, $\cos 2\theta$, $\sin 4\theta$, and $\cos 4\theta$. The measured torque curves (Fig. 3) were found to be composed of terms proportional to $\sin 2\theta$, $\cos 2\theta$, $\sin 4\theta$, and $\cos 4\theta$, with no appreciable residual. Accordingly, the curves were analyzed into these components,[3] the calculations being facilitated by the use of a punched card machine. Writing the expression for L in the form

$$L=a_2\sin 2\theta+a_4\sin 4\theta+b_2\cos 2\theta+b_4\cos 4\theta, \quad (3)$$

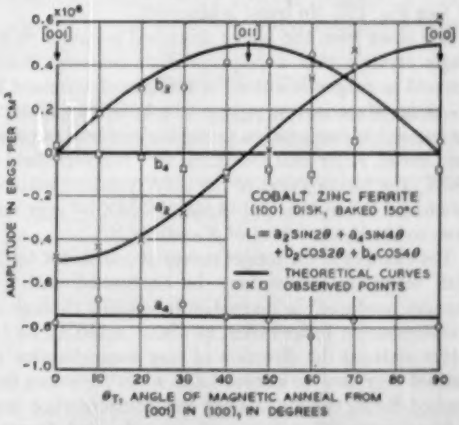


FIG. 4. Amplitudes of 2θ and 4θ components of torque curves [see Eq. (3)].

[3] This was done with the kind cooperation of Dr. R. W. Hamming and Miss F. Maier.

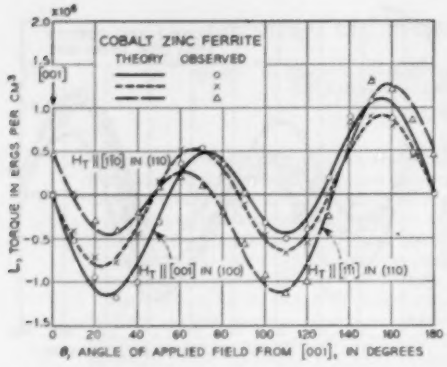

COBALT ZINC FERRITE
THEORY OBSERVED
[001]
$H_T \parallel [\bar{1}10]$ IN (110)
$H_T \parallel [001]$ IN (100)
$H_T \parallel [111]$ IN (110)
L_2, TORQUE IN ERGS PER CM³
θ, ANGLE OF APPLIED FIELD FROM [001], IN DEGREES

FIG. 5. Observed points, and lines calculated for three principal cases, using constants $K_1=1.5\times10^6$, $K_T=0.5\times10^6$.

and comparing with the previous expression for L, we have

$$a_2=-K_T\cos2\theta_T, \quad a_4=-K_1/2, \quad b_2=K_T\sin2\theta_T, \quad b_4=0.$$

The values of a_2 to b_4, derived by analysis of the torque curves, are plotted in Fig. 4 as points and there they can be compared with the expected values shown by the lines drawn for $K_1=1.5\times10^6$, $K_T=0.5\times10^6$ ergs/cm³.

It will be noted in Fig. 3 that the direction of easy magnetization may depart from the cubic axis by somewhat more than 10°. Larger departures are noted in Fig. 5; here points show measurements made on the (100) and (110) specimens and the curves are calculated from Eq. (1) using the above chosen values of K_1 and K_T. Theoretical and observed directions of easy magnetization for the (100) disk are plotted in Fig. 6 as lines and points, respectively.

To test for any possible variation of K_T with θ_T, K_T was derived from each curve of Fig. 3 and plotted against θ_T (see Fig. 19). No trend is observed.

It follows from the above described behavior of the single crystals that a polycrystalline material should respond to magnetic anneal in a degree determined by K_T alone. Material was prepared by L. G. Van Uitert to correspond in composition to the single crystal of cobalt zinc ferrite. After heat treatment in a magnetic field at 150°C, the torque curve of the polycrystalline material was found to have an amplitude of 0.54×10^6 ergs/cm³, close to the chosen value of $K_T=0.5\times10^6$.

The changes in the torque curves of cobalt zinc ferrite with magnetic anneal may be contrasted with the changes produced[4] in magnetite by cooling through its transformation temperature at about −140°C. In the latter material the direction of easy magnetization remained very close to a crystal axis when the strong field applied during cooling through the transformation temperature was 30° or less from the axis, and departed widely from an axis when the field was near 45°. This change of the direction of easy magnetization with the direction of the field present during cooling is shown for magnetite in Fig. 6, curve E, where it can be compared with the data for cobalt zinc ferrite, curve A. The different behavior of the two materials suggests that different mechanisms are operating.

Two other estimates have been made of the anisotropy of cobalt ferrite. Weisz[5] has derived the absolute value of K_1 from measurements of the initial permability, μ_0, of polycrystalline material, using the relation

$$\mu_0-1=21M_s^2/K_1.$$

This assumes that the change in magnetization takes place by a rotational process alone, and assumes a partly empirical constant of proportionality. His value of $|K_1|=3.5\times10^6$ compares with our measured value of $K_1=1.8\times10^6$, for the material nearest stoichiometry. Constants for other stoichiometric simple ferrites are given in his original paper. In the case of stoichiometric $NiFe_2O_4$ Weisz's estimated value of $|K_1|=400\,000$ is to be compared with $K_1=62\,000$ determined by Galt, Matthias and Remeika.[6]

Shenker[7] has measured K for cobalt and nickel ferrites by observing the torque in a measured field which is insufficient to saturate. Results are reported for temperatures of 20 to 800°K. The method is based essentially on the known relations

$$L=-HM_s\sin(\theta_0-\theta)=-\tfrac{1}{2}K_1\sin4\theta,$$

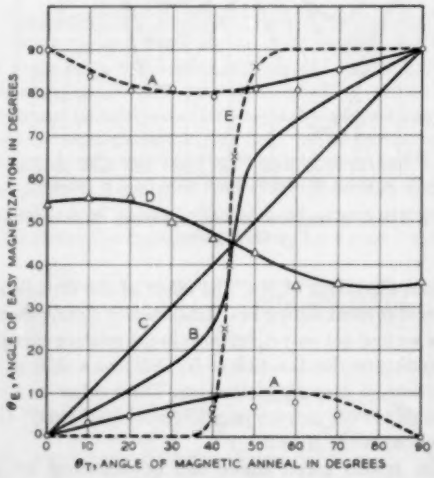


FIG. 6. Dependence of direction of easy magnetization on the direction in which the field was applied during heat treatment in the (100) plane. Points, observations. Lines are: (A) for chosen values $K_T/K=0.5/1.5=\frac{1}{3}$, (B) for $K_T/K=1$, (C) for $K_T/K=\infty$. Theoretical and observed directions of hard magnetization are shown by (D). Data for Fe_3O_4 are represented by broken line (E), to show the difference in type of curves. In (A), solid lines show easiest directions, dotted lines, secondary easy directions.

[4] Williams, Bozorth, and Goertz, J. Appl. Phys. **91**, 1107 (1953).

[5] R. S. Weisz, Phys. Rev. **96**, 800 (1954).

[6] Galt, Matthias, and Remeika, Phys. Rev. **79**, 391 (1950).

[7] H. Shenker, thesis, University of Maryland, 1955 (unpublished); and private communication.

θ and θ_0 being the respective angles that M_s and H make with [001] in the (100) plane. When θ_0 is small, L can be plotted against $1/H$ and extrapolated to $1/H=0$. This gives the torque for infinite field, which is $-2K\theta_0$, and so K_1 is evaluated. M_s may also be obtained from the slope of the curve used for extrapolation.

A more exact method is to plot a series of curves for $2/\sin 4\theta$ *vs* $\sin(\theta_0-\theta)$, for various values of θ_0. Then for given θ_0 one measures $-L/HM_s$ [equal to $\sin(\theta_0-\theta)$] and reads off $-K/L$ (equal to $2/\sin 4\theta$) from the appropriate curve, and then obtains K_1 from the measured value of L. This assumes that M_s is known.

Shenker's values of K for a crystal of composition $Co_{1.12}Fe_{2.21}O_4$ are 3.8×10^6 for 20°C and 17.5×10^6 for −196°C. The former may be compared (Table I) with our values for specimens of different compositions. Shenker's values of K_1 for $Ni_{0.47}Co_{0.06}Fe_{1.39}O_4$ for these temperatures are also given in this table ($K_1=-42\,900$ and $-27\,800$, respectively).

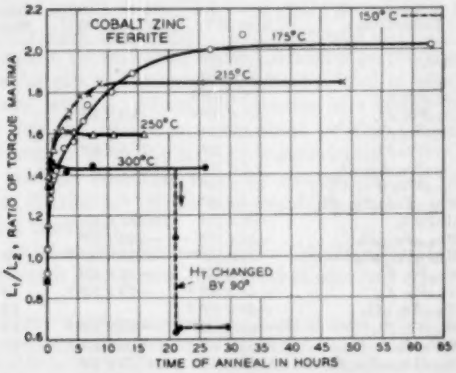


FIG. 7. Anisotropy as dependent on time of magnetic anneal. L_1 and L_2 as designated in Fig. 1, curve *B*. Dotted curve shows effect of changing direction of magnetic anneal by 90 degrees, from [001] to [010].

TEMPERATURE AND TIME EFFECTS

The measurements described above were made after the specimen had been held at 150°C for three days. The magnitude of the effect is dependent on both the temperature and time of treatment in a field, as well as on the magnitude of the field. As a measure of the effect we used the ratio of the first minimum to the first maximum of the torque curves, designated as in Fig. 1 as L_1 and L_2. This ratio is plotted in Fig. 7. The change of L_1/L_2 with time of treatment shows that a change is taking place, to which a relaxation time may be assigned. Plotting the logarithm of this relaxation time against the reciprocal of the absolute temperature, as shown in Fig. 8, one obtains the activation energy of 0.94 ev.

It is apparent from Fig. 7 that a different limiting value of L_1/L_2 is approached for each temperature of magnetic anneal. The limiting values are plotted against

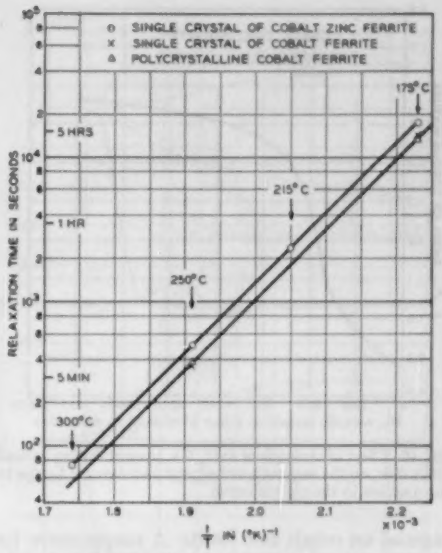


FIG. 8. Relaxation time of effect of magnetic anneal for (a) cobalt ferrite single crystal, (b) cobalt zinc ferrite single crystal, (c) cobalt ferrite polycrystal. Activation energy, 0.94 ev.

temperature in Fig. 9. The ratio is higher the lower the temperature of anneal, and drops in an almost linear fashion to one at the Curie point.

The same specimen, a disk with dimensional ratio 2.36, was used to determine the strength of the field necessary for effecting a magnetic anneal. The data of Fig. 10 show that an applied field of 500 oersteds will cause about half of the maximum change in L_1/L_2. Since this is about half of the demagnetizing field for saturation magnetization, the field inside the material (applied field minus demagnetizing field) is presumably much smaller. This is confirmed by measurements on a hollow square specimen as described below.

OTHER FERRITES

The single crystal of cobalt ferrite of composition $Co_{0.78}Fe_{2.27}O_4$ was used in some experiments like those

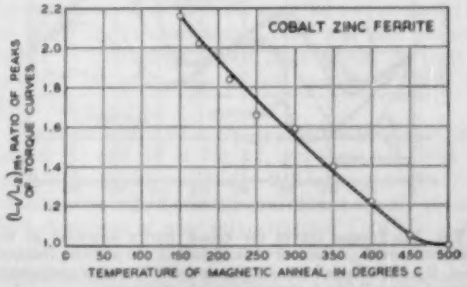


FIG. 9. Effect of magnetic anneal at various temperatures, measured by the ratio of the two peaks of the torque curves in the (100) plane.

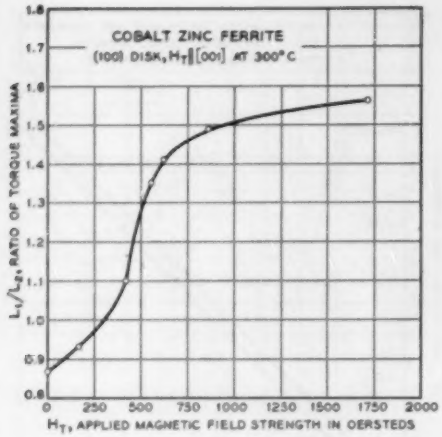


FIG. 10. Effect of strength of field, H_T, present during annealing of (100) disk, on the magnetic anisotropy as measured by the ratio of the maxima in the torque curve.

performed on cobalt zinc ferrite. A torque curve for a (100) disk, heated at 150°C with $H_T \| [001]$, is shown by the solid line in Fig. 11. Before analyzing into its 2θ and 4θ components this curve had to be corrected[8] for lack of saturation even in a field strength of 25 000 oersteds. The corrected points are shown as triangles, and the accompanying dotted curve is the best line through them that contains only sines and cosines of 2θ and 4θ. The analysis gives $K_1 = 4.3\times10^6$, $K_T = 2.5\times10^6$ ergs/cm^3, values considerably higher than those for cobalt zinc ferrite. The curve obtained after treatment in zero field was somewhat irregular in shape, showing definite lack of saturation and possibly other disturbing effects. Therefore K_1 was deduced from the curve for $H_T \| [001]$, as mentioned above.

The cobalt ferrite of composition $Co_{1.1}Fe_{1.9}O_4$, nearly

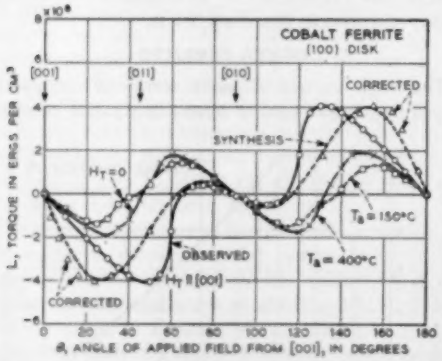


FIG. 11. Torque curves for cobalt ferrite annealed at two temperatures, T_a, with and without presence of a strong magnetic field, H_T. "Synthesis" curve is drawn from 2θ and 4θ components derived originally from corrected points (triangles).

[8] As proposed by H. J. Williams. See Bozorth, Phys. Rev. **96**, 311 (1954).

stoichiometric, was found to have the much lower crystal anisotropy constant of $K_1 = 1.8\times10^6$. It did not respond to magnetic anneal, therefore $K_T = 0$.

Some other ferrites, of manganese, nickel, and manganese-zinc, were examined and the results are given in Table I.[4,7,9–15] Direct comparison was made with the results on manganese ferrites obtained at microwave frequencies.[9] Pieces for both kinds of experiments were cut from the same ferrite crystal. It is noted that the results are the same for the two methods, well within the experimental error.

The results for nickel ferrite at room temperature may be compared with the previous data obtained by Galt and collaborators[16,17] and by Healy.[11] Some differences are apparent. Not enough data are available to determine quantitatively the effect of small differences in chemical composition.

TABLE I. Crystal anisotropy constants of some ferrites.

Composition	K_1 at 20°C	K_1 at −196°C	Reference
Fe_3O_4	-1.1×10^6	...	12
Fe_3O_4	-1.35×10^6	...	4
[b] $Co_{0.8}Fe_{2.2}O_4$	3.9×10^6 [a]		
[b] $Co_{0.8}Fe_{2.2}O_4$	2.9×10^6 [c]	4.4×10^6	
$Co_{1.1}Fe_{1.9}O_4$	1.8×10^6		
$Co_{1.1}Fe_{1.9}O_4$	3.8×10^6	17.5×10^6	7
$Co_{0.7}Zn_{0.3}Fe_{2.2}O_4$	1.5×10^6		
$Co_{0.3}Mn_{0.4}Fe_{2.3}O_4$	1.1×10^6		
$Mn_{0.45}Zn_{0.55}Fe_{2.0}O_4$	-3.8×10^3		13
$MnFeO_4$	-28×10^3	-187×10^3	
[b] $Mn_{0.98}Fe_{1.88}O_4$	-34×10^3	-240×10^3	
[b] $Mn_{0.98}Fe_{1.88}O_4$	-33×10^3 [d]	-233×10^3 [d]	9
[b] $Ni_{0.8}Fe_{2.2}O_4$	-39×10^3	-42×10^3	
[b] $Ni_{0.8}Fe_{2.2}O_4$	-43×10^3 [d]	-74×10^3 [d]	10
$Ni_{0.75}Fe_{2.25}O_4$	-49×10^3 [d]		14
$Ni_{1.8}Fe_{2.0}O_4$	-63×10^3 [d]	-87×10^3 [d]	15
$NiFe_2O_4$	-51×10^3 [d]	$+20\times10^3$ [d]	11
$Ni_{0.7}Co_{0.003}Fe_{2.3}O_4$	-18×10^3	-125×10^3	
$Ni_{0.7}Co_{0.004}Fe_{2.3}O_4$	-10×10^3	-196×10^3	
$Ni_{0.47}Co_{0.03}Fe_{1.4}O_4$	-43×10^3	-28×10^3	7

[a] Quenched in air from 400°C.
[b] Brackets indicate that measurements were made on different specimens cut from the same crystal.
[c] Aged at 150°C for 3 days.
[d] Constant derived from measurements at microwave frequencies. Others at zero frequency (torque method).

Bickford *et al.*[18] have shown that addition of only a small amount of cobalt ferrite to magnetite changes the sign of the anisotropy constant from negative to positive. Our experiments extend this idea to additions to nickel ferrite. The anisotropy constant of nickel ferrite is reduced from about −40 000 to −10 000 by the

[9] Private communication from S. Geschwind and J. F. Dillon.
[10] Private communication from Cetlin, Galt, Merritt, and Yager.
[11] D. W. Healy, Phys. Rev. **86**, 1009 (1952).
[12] L. R. Brickford, Phys. Rev. **78**, 449 (1950).
[13] Galt, Yager, Remeika, and Merritt, Phys. Rev. **81**, 470 (1951).
[14] Yager, Galt, and Merritt, Phys. Rev. **99**, 1203 (1955).
[15] Yager, Galt, Merritt, and Wood, Phys. Rev. **80**, 774 (1950).
[16] Galt, Yager, and Merritt, Phys. Rev. **93**, 1119 (1954).
[17] J. K. Galt, Bell System Tech. J. **33**, 1023 (1954).
[18] Bickford, Pappis, and Stull, Office of Naval Research Technical Report, January, 1954 (unpublished).

addition of only 0.4 mole percent of cobalt ferrite (attempted composition).

It is possible that the variable estimates of K_1 for nickel ferrite may be attributed to a slight variable impurity of cobalt.

Specimens cut from the same crystal of $Ni_{0.8}Fe_{2.2}O_4$ were used for microwave measurements by Cetlin, Galt, Merritt, and Yager, and for static measurements by the authors. Results show a decided difference between the K_1's determined at −196°C, and some difference also at room temperature. The large difference at −196°C is in accord with the relaxation phenomena reported by Galt *et al.*[16,17] This is discussed in a recent note.[18a]

MAGNETOSTRICTION

Measurements of magnetostriction and of crystal anisotropy were made on the same crystals. Using the strain-gauge technique,[19] magnetostriction was determined usually in the [001] and [011] directions in the (100) plane, as dependent on field strength up to 25 000 oersteds, the field being applied parallel or perpendicular to the gauge. Specimens were heat-treated in various ways, with or without the presence of a strong field during the heat treatment. Some of the materials showed an unusually large magnetostriction.

One can determine from the magnetostriction data the distribution of domains resulting from a given heat treatment.

Results are shown in Fig. 12 for $Co_{0.2}Zn_{0.2}Fe_{2.2}O_4$, the specimen having been treated (a) in zero field ($H_T=0$) and (b) in a high field parallel to [001]. For this and other substances examined the results are interpreted in terms of the coefficients λ_{100} and λ_{111} which appear in the well-known expression[20] for the magnetostriction $\lambda=\Delta l/l$ in a cubic crystal:

$$\lambda=(3/2)\lambda_{100}(\alpha_1^2\beta_1^2+\alpha_2^2\beta_2^2+\alpha_3^2\beta_3^2-1/3)+3\lambda_{111}(\alpha_1\alpha_2\beta_1\beta_2+\alpha_2\alpha_3\beta_2\beta_3+\alpha_3\alpha_1\beta_3\beta_1).$$

TABLE II. Magnetostriction constants of some ferrites. Also saturation magneostriction, $\bar{\lambda}$, of polycrystalline material, calculated from these constants. Values are for room temperature unless noted.

Composition	$\lambda_{100}\times10^6$	$\lambda_{111}\times10^6$	$\bar{\lambda}\times10$
Fe_3O_4[19]	−19	81	41
Fe_3O_4(124°K)[18]	−23	55	24
$Co_{1.1}Fe_{1.9}O_4$	−250	—	—
$Co_{0.8}Fe_{2.2}O_4$	−590	120	−210
$Co_{0.2}Zn_{0.2}Fe_{2.2}O_4$	−210	110	−18
$Co_{0.2}Mn_{0.4}Fe_{2.4}O_4$	−200	65	−40
$Ni_{0.8}Fe_{2.2}O_4$	−36	−4	−17
$Mn_{0.88}Fe_{1.88}O_4$	−35	−1	−15
$Mn_{0.4}Zn_{0.4}Fe_{2.2}O_4$	−14	14	3

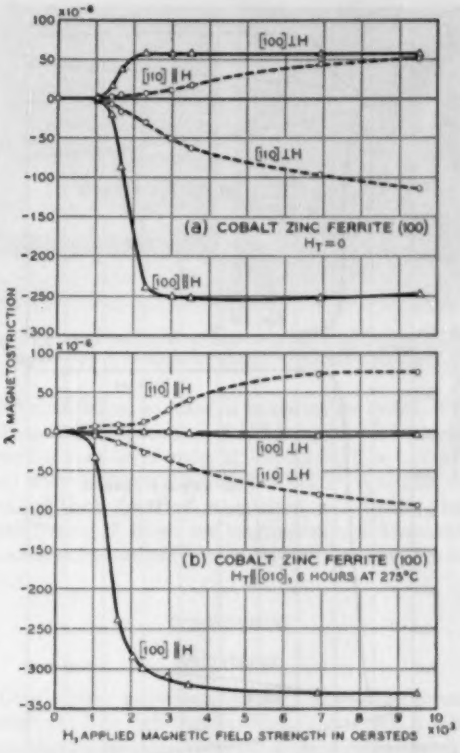


FIG. 12. Magnetostriction in Co-Zn-ferrite after treatment in (a) zero field, and (b) strong field parallel to [010] in (100) plane.

Here the α's are the direction cosines of the saturation magnetization, and the β's those of the measured change in length, with respect to the crystal axes. In any given specimen, with given distribution of domains in the initial (demagnetized) state, the constants are evaluated by magnetizing to saturation first in one direction, then in another, and noting the difference. In most of the experiments reported here, λ_{100} was determined by placing a gauge parallel to [001] in (100) and measuring the change in length when the material was saturated in the plane in directions first parallel and then perpendicular to the gauge. Then $\lambda_{100}=\lambda_{\parallel}-\lambda_{\perp}$ in [001]; similarly, using a gauge placed parallel to [011] in (100), $\lambda_{111}=\lambda_{\parallel}-\lambda_{\perp}$.

The same values of λ_{100} and λ_{111} were derived from the data of (a) and (b) of Fig. 12, and these are given in Table II. In (a), for which $H_T=0$, it is apparent that the domains were initially distributed almost equally among the 6 directions of easy magnetization, [001], ···. In (b), for which $H_T\|[001]$, the domains were initially aligned parallel and antiparallel to [001], where they were fixed by the magnetic anneal. It should be noted that the magnetostriction in this direction is zero; therefore, essentially all of the domains were oriented

[18a] Bozorth, Cetlin, Galt, Merritt, and Yager, Phys. Rev. 99, 1898 (1955).
[19] J. E. Goldman, Phys. Rev. 72, 529 (1947).
[20] See reference 2, p. 650.

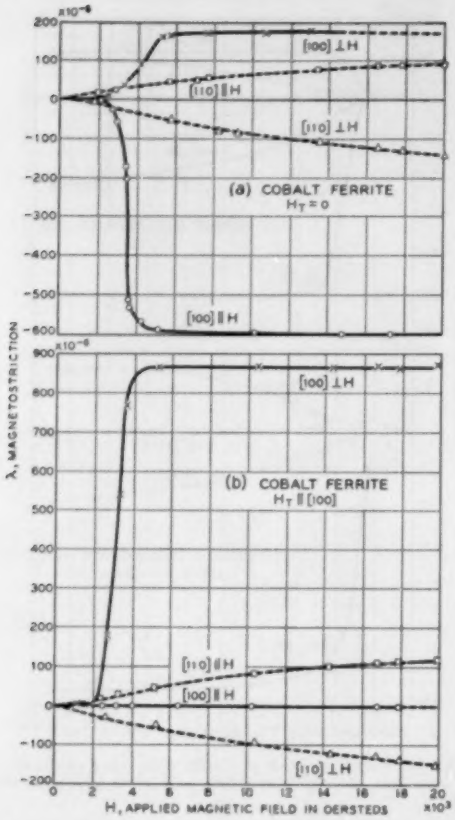


FIG. 13. Magnetostriction of cobalt ferrite crystal (a) without and (b) with magnetic anneal. Derived constants in Table II.

originally parallel to [001], although a shallow minimum in the crystal anisotropy energy occurs also ∥[010], a subsidiary direction of easy magnetization as shown by the torque curve *B* of Fig. 2.

Similar data for $Co_{0.8}Fe_{2.2}O_4$ are shown in Figs. 13 (a) and (b), and the derived values of λ_{100} and λ_{111} are the same for the two sets (see Table II). The initial domain distribution is fixed, as for cobalt zinc ferrite, by the magnetic anneal (direction of H_T), when this is parallel to a crystal axis.

For this material the magnetostriction measured with the field and gauge perpendicular to H_T is very large, namely 740×10^{-6}. Measurements of this were repeated with a second specimen comprised of a rod of dimensional ratio length/thickness = 2.2 instead of with a disk. See Fig. 14. It was noted that the λ *vs* H curve is strictly reversible, even on the steep portion, the same points *A* and *B* being obtained on increasing as on decreasing H. The slope on the steep portion is equal to $d\lambda/dH=0.5\times10^{-6}$, and would presumably be much higher if the demagnetizing field of the rod (about 1000 oersteds at saturation) were smaller.

The smallness of the initial slope of the λ *vs* H curve is apparently associated with the high anisotropy of the specimen. The corresponding curve for the specimen heat-treated in zero field has a larger initial slope because in this material 90° domain walls can move so as to permit the magnetization to change from one direction of easy magnetization to another equally easy. When H_T is applied parallel to one crystal axis, the difference in the crystal anisotropy energy along the axis and one at right angles is $K_T=2\times10^6$ ergs/cm³.

The magnetostriction constants of single crystals of these and some other ferrites are given in Table II. The saturation magnetostriction of polycrystalline ferrites of these compositions are included in the table, as calculated from the relation $\bar{\lambda}=(2\lambda_{100}+3\lambda_{111})/5$.

HYSTERESIS LOOPS

Since materials that respond to magnetic anneal often have square hysteresis loops, a specimen was cut from a single crystal in the form of a hollow square so that it could be measured with closed magnetic circuit. The length of the edges, cut parallel to [100] directions, was 3.7 mm, the thickness 1.2 mm, and the width of each leg 1.1 mm. Hysteresis loops for different fields present during annealing for 15 minutes at 400°C are shown in Fig. 15. Although sharp corners were not observed on the loops, remanence was considerably increased by magnetic anneal and the maximum effect was obtained when the field present during anneal, H_T, was about 10 oersteds. The coercive force was then about 5 oersteds. The lack of sharp corners may be due to the geometrical imperfection of the specimen. Unfortunately, the speci-

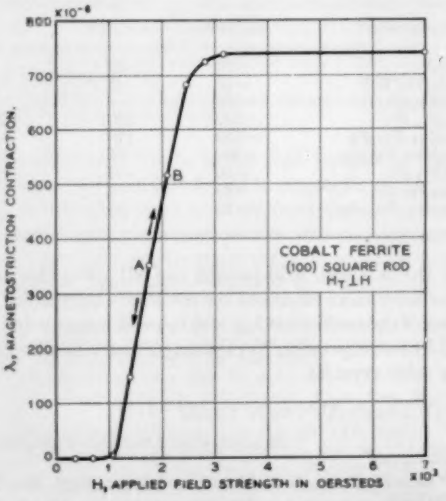


FIG. 14. Magnetostriction of cobalt ferrite square rod. Note reversibility on the steep portion of the curve.

men was broken before it could be annealed at right angles to its plane.

Powder patterns were made from portions of the specimen by Mr. R. C. Sherwood, after careful polishing and etching. These show the expected domain structure with parallel and antiparallel domains, the domain walls being always parallel to the edges of the magnetization and to the specimen. Figure 16 shows two of the patterns for the demagnetized state.

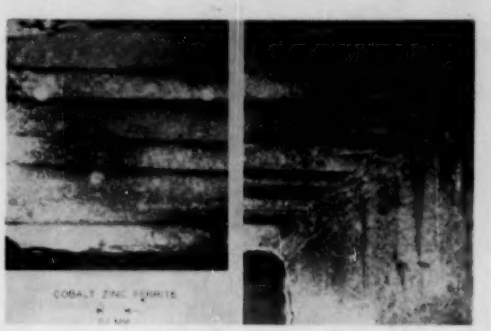


FIG. 16. Powder pattern on hollow rectangle of cobalt zinc ferrite, showing effect of magnetic anneal on domain orientation.

POLYCRYSTALLINE SPECIMENS

The experiments on single crystals of cobalt ferrite showed that the response of the material to magnetic anneal varied with composition. This response could be investigated most easily with polycrystalline specimens, and was determined for several compositions lying between $CoFe_2O_4$ and Fe_3O_4. Specimens were fired at 1250°C in O_2 or CO_2, depending on composition. The magnetic anisotropy created by the magnetic anneal, expressed as maximum torque per unit volume and given in Fig. 17, shows the already known lack of response at the two extremes and a rather high maximum in the middle. The highest measured anisotropy corresponds to $K_T = 2\times10^4$ ergs/cm³.

The composition for maximum effect of heat treatment is consistent with the optimum composition of material used for permanent magnets, known as "O.P." or "Vectolite."

Ordinarily the magnetostriction of a material at saturation is considered to be a structure-insensitive property. In view of the change in some of the properties of ferrites with state of oxidation, it was suggested by L. G. Van Uitert that the dependence of saturation magnetostriction on state of oxidation be tested. Two polycrystalline specimens of nickel ferrite were selected from the same lot of material, having Fe/Ni = 1.87, and fired under different conditions of oxidation so that the ratio of their electrical resistivities was greater than 1000. Figure 18 shows the longitudinal and transverse magnetostrictions to be the same, within experimental error.

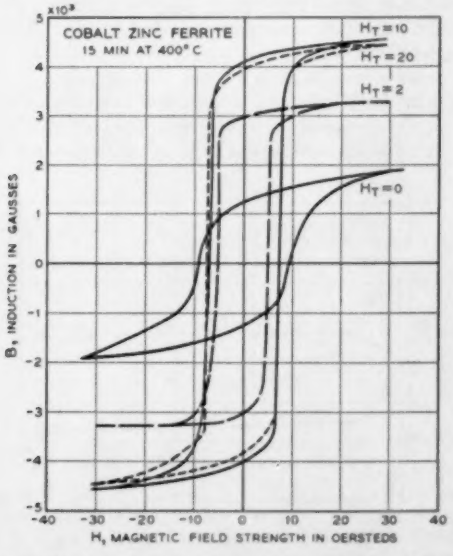


FIG. 15. Hysteresis loops of cobalt zinc ferrite showing effect of magnitude of field present during anneal, H_T, in oersteds.

DISCUSSION

Anisotropy

Cobalt ferrite is characterized by high-positive crystal anisotropy, by a high-negative magnetostriction, and by response to magnetic anneal. All of these characteristics are at a maximum when some iron ferrite (magnetite) is present in solid solution, and they tend to lower values (and zero response to magnetic anneal) as the composition approaches stoichiometric $CoFe_2O_4$ or Fe_3O_4.

Addition of a small amount of cobalt ferrite to other ferrites is often enough to change their inherent negative anisotropy (easy direction, [111]) to positive,[21] and to make them respond to magnetic anneal. Addition of about 1 percent of $CoFe_2O_4$ to iron-rich $NiFe_2O_4$ seems to be enough to make the negative anisotropy of the latter decrease in magnitude and pass through zero to become positive.

Since the magnetostriction is large there is some possibility that it will contribute substantially to the anisotropy. The contribution has been expressed by Becker and Döring[22] as an additional anisotropy constant, ΔK_1:

$$K_1 = K_0 + \Delta K_1 = K_0 + (9/4)[\lambda_{100}(C_{11} - C_{12}) - \lambda_{111}C_{44}],$$

K_0 being the intrinsic crystal anisotropy constant, K_1 the observed constant, and ΔK the difference which is dependent on the magnetostriction and elastic constants. Calculation of $\Delta K_1/K_1$ for cobalt zinc ferrite, using the elastic constants determined by McSkimin

[21] J. L. Stull and L. R. Bickford, Phys. Rev. 92, 845 (1953).
[22] R. Becker and W. Döring, *Ferromagnetismus* (Verlag Julius Springer, Berlin, 1939), p. 145.

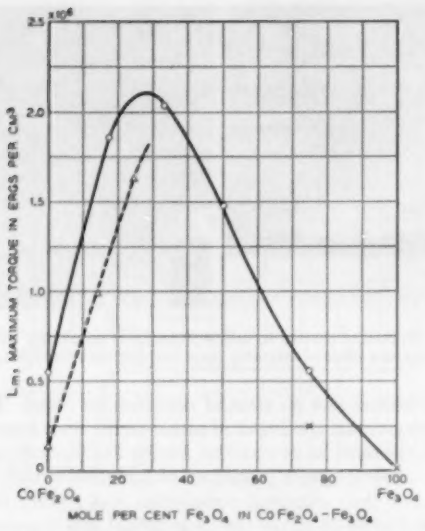


FIG. 17. Dependence of anisotropy, developed in polycrystal line material by heat treatment, on composition. Maximum torque, L_m, is equal to $K_T/2$. Attempted compositions; circles, fired in oxygen; crosses, fired in CO_2; triangles, another series fired in air.

and others,[22] is found to be only $70\times10^3/(1.5\times10^6)$ or less than 5 percent. Thus the contribution of magnetostriction to the anisotropy is small.

In the crystal of cobalt zinc ferrite the response to magnetic anneal was such as to make the selected direction an axis of easier magnetization, the difference in the energy between this direction and the direction at right angles being 0.5×10^6 ergs/cm³. In iron-rich cobalt ferrite the response is similar in kind, but larger.

The origin of this difference in energy is so far not clear. The firmness by which the domains are held in the easy direction by the magnetic anneal, measured by K_T, can be compared with the elastic energy that is absorbed during the change of magnetization from one easy direction to another. Consider a crystal that has been heat-treated with the field parallel to [100], so that all domains are parallel to this axis, and then magnetize it to saturation parallel to another axis, [010]. The elastic energy so expended is λ^2Y, where $\lambda=\frac{3}{2}\lambda_{100}$ and Y represents Young's modulus, and is

$$E=(9/4)\lambda_{100}(c_{11}-c_{12})(c_{11}+2c_{12})/(c_{11}+c_{12}).$$

This is 150 000 ergs/cm³, which is to be compared with 500 000 ergs/cm³, the difference in energy as determined from the torque curve. The domains are obviously held to the direction of easy magnetization by forces much stronger than those of magnetostriction.

Néel[24] has proposed a theory of short range order and its effect on the magnetostriction and crystal anisotropy. In relation to the present experiments it suggests that the short range order is controlled by the magnetic anneal and that as a result there will be a uniaxial anisotropy superposed on the normal crystal anisotropy. The magnitude of the uniaxial anisotropy, however, should be dependent on the crystallographic direction according to Néel's theory, and in a face-centered cubic lattice the constant K_T should be twice as large when $H_T\|[110]$ as it is when $H_T\|[100]$. In our experiments no such crystallographic dependence was observed. In order to check on this point our torque curves were analyzed and K_T, the amplitude of the 2θ component, was derived separately from each curve for the (001) plane. The results, given in Fig. 19, show no trend that approaches the magnitude required by Néel's theory for a *simple* face-centered lattice. In a private communication Professor Néel has pointed out that the various interactions between neighbors, either between nearest-neighbors, second-nearest neighbors, or through super-exchange, is so complicated that it is uncertain if there should be any dependence of K_T on crystallographic direction in the ferrites. The results, then, are not necessarily in conflict with Néel's theory.

Note added in proof.—In an attempt to detect atomic ordering, E. Prince has made neutron diffraction observations on cobalt ferrites of various compositions, both before and after heat treatment in a magnetic field. The neutron data show no evidence of long range order of Fe and Co on octahedral sites, but the possibility of some short range order cannot be eliminated. The magnetic structures of heat-treated samples contain moments which are not in general aligned parallel to [100] directions, the angle of misalignment being greater in iron-rich samples in accordance with the magnetic observations. We are indebted to Dr. Prince for permission to include this note in our paper.

Chikazumi[25] measured the effect of magnetic anneal on the anisotropy of a crystal of composition $FeNi_3$, and found a uniaxial anisotropy, K_T, of magnitude 450 to

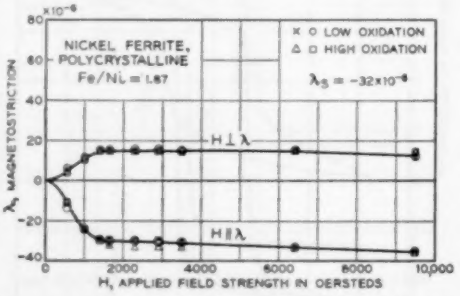


FIG. 18. Magnetostriction of polycrystalline nickel ferrite, showing negligible effect of oxidizing anneal which changed conductivity 1000-fold.

[22] Using the elastic constants determined on a piece of the same crystal by McSkimin, Williams, and Bozorth, Phys. Rev. **95**, 616 (1954).

[24] L. Néel, J. phys. radium [8] **15**, 225 (1954).

[25] S. Chikazumi, as reported by S. Kaya, Revs. Modern Phys. **25**, 49 (1953).

1250 ergs/cm³. In accordance with Néel's theory, and unlike our measurements on Co-Zn-ferrite, the uniaxial anisotropy of $FeNi_3$ depends on the crystallographic direction.

Another possible explanation to be considered in the explanation of our results is the formation of thin plates or rods of precipitate that are directed by the magnetic field during the anneal, in much the way that they are[26] in Fe_2NiAl and Alnico 5. According to calculation, the torque of a rod-like precipitate at saturation is

$$L=\tfrac{1}{2}(N_t-N_l)M_s^2\sin 2\theta,$$

N_l and N_t being the longitudinal and transverse demagnetizing factors of the rods. In cobalt zinc ferrite the amplitude of the 2θ term, 0.5×10^6, would be realized by a rod-like precipitate having $N_t-N_l\approx 2\pi$, corresponding to a large ratio of length/diameter. Plates having appropriate dimensional ratios might also satisfy the experimental data. The nonsinusoidal character of the torque curves of cobalt ferrite, seen in Fig. 11, supports the idea that a separate phase is precipitated at low temperatures and that its direction is controlled by the field during annealing.

The time constants used in determining the energy of activation, (Figs. 7 and 8), correspond to the expression for the relaxation time:

$$\tau=\tau_0 e^{w/kT},$$

the energy of activation being 0.94 ev. The magnitude is consistent with the time necessary for the diffusion of atoms in a solid of this kind. No specific data on the diffusion constants of any of the elements in these materials, or their temperature dependence, are known to us.

It is surprising at first to find the relaxation time of magnetic anneal, as shown in Fig. 7, to be so rapid at temperatures as low as 150°C. It should be remembered, however, that the ferrite structure is built up of a lattice of relatively large and deformable 0^{--} ions with much smaller metal ions in the interstices. It is therefore reasonable to observe the effects of the diffusion of the relatively small metal atoms at this low a temperature; the diffusion of C and N in iron take place with comparable speeds at even lower temperatures. The relaxation time of cobalt ferrite reduces to about 1 second at the Curie point.

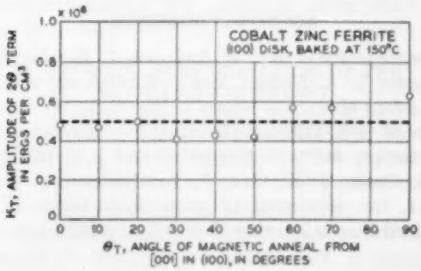


FIG. 19. Constants K_T derived from harmonic analysis of torque measured in (100) plane, for various angles θ_T between H and [001].

[26] Nesbitt, Williams, and Bozorth, J. Appl. Phys. 25, 1014 (1954).

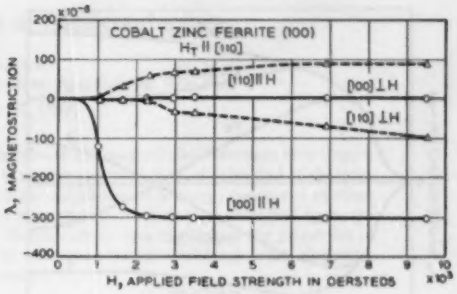


FIG. 20. Magnetostriction of Co–Zn-ferrite in (100) plane. See Table II for derived constants.

The temperatures required for the sintering of ferrite powder are hundreds of degrees higher than the temperatures at which magnetic anneal is effective. It is suggested that the diffusion of the 0^{--} ions, which control the lattice structure, occurs with appreciable rapidity only at the higher temperatures needed for sintering, while the much greater mobility of the metal ions at the lower temperatures controls the magnetic properties.

It will be noted that τ_0 for the process under discussion is about 10^{-6} sec, a time long compared with 10^{-13} sec, the time of oscillation of an atom in the lattice. This means that many oscillations take place before an atom can escape from a potential well, or that it must escape successively from many wells before it finds a position where it causes a change in the property observed, in this case crystal anisotropy.

Magnetostriction

Composition.—Insufficient data are available to show uniquely the effect of composition on the magnetostriction constants. It is apparent, however, that the presence of Co causes a large negative value of λ_{100} and a moderately large positive λ_{111}. Apparently higher λ_{100}'s are obtained when there is somewhat less than one Co atom per molecule. The effect of the valence of the atoms (in nickel ferrite) has no apparent effect on magnetostriction.

Crystal symmetry.—In the ferrites examined $\lambda_{100}<0$, and generally $\lambda_{111}>0$. In most cases $|\lambda_{100}|>|\lambda_{111}|$, and in some cases $|\lambda_{100}|$ is very large. The high magnetostriction emphasizes the fact that crystals which are originally cubic have lower symmetry when magnetized. Guillaud and Sage[27] have even observed by x-rays the difference in spacing between (001) and (100) planes

[27] C. Guillaud and M. Sage, Compt. rend. 237, 313 (1953). See also C. Guillaud, Revs. Modern Phys. 25, 64 (1953).

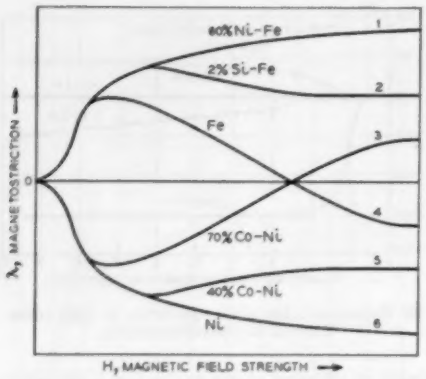


FIG. 21. Various types of magnetostriction curves expected from various relations between K, λ_{100} and λ_{111}. See Table III.

when the direction of magnetization is differently related to these planes.

This raises the question as to when to take account of the effect of the magnetostriction on the crystal symmetry. One point of view that we believe to be tenable is the following. When a crystal such as iron is unmagnetized and its spins are distributed equally among the six directions of easy magnetization, it is cubic, because the properties measured in various directions in a crystal of reasonable size have the required symmetry (four 3-fold symmetry axes). This is so even though the unit cell, and also a single domain, have only tetragonal symmetry, and these subdivisions of nickel have rhombohedral symmetry. If either of these crystals is magnetized along one axis it is tetragonal, and if magnetized in an arbitrary direction it is triclinic.

Initial domain distribution.—The magnetostriction at saturation is useful not only in determining the values of the constants λ_{100} and λ_{111} but also in deducing the domain distribution in the unmagnetized state. For example, Fig. 13 shows that when cobalt ferrite is magnetically annealed with $H_T \| [100]$, the longitudinal magnetostriction in this direction is zero, therefore in the unmagnetized state all of the domains are oriented either parallel or antiparallel to this direction. Even though [010] is also a direction of easy magnetization, practically all of the domains are fixed $\| [100]$.

After annealing the same specimen with $H_T \| [110]$ it was found (see Fig. 20) that practically all of the domains were aligned nearly parallel to [100] and none $\| [010]$. In this case [100] and [010] are equally easy and equally near to H_T, and one might expect half of the domains to be nearly parallel to each axis. Presumably H_T was slightly nearer to one axis than to the other and the nearer axis was chosen. It may be that a very slight error in adjusting the angle of H_T will result in the alignment of the domains almost parallel to one or the other of the axes. The torque curve for this treatment (see Fig. 6) shows the easy direction to lie about 10° from the crystal axes.

Form of λ, H *curve.*—Bickford, Pappis, and Stull[18] have studied the behavior of λ *vs* H curves of Fe_3O_4 and of solid solutions containing a few percent of $CoFe_2O_4$. They note that in Fe_3O_4 at room temperature the magnetostriction is positive at low values of H, and that it becomes negative when the temperature is lowered below 160°K and [100] becomes the direction of easy magnetization ($K_1>0$). As they point out, the initial part of the λ, H curve depends on the sign of λ in the easy direction, and the latter part of the curve shows the effect of the sign of λ in the hard direction.

This idea may be generalized, and the relation of the signs of K_1, λ_{100}, and λ_{111} to the general shape of the λ, H curve may be tabulated. It is convenient to distinguish

TABLE III. Types of magnetostriction curves (λ *vs* H). See Fig. 20.

K_1	λ_{100}	λ_{111}	Curve type	Examples
+	+	+	1	60% Ni–Fe
+	+	−	2, 4	2% Si–Fe, Fe
+	−	+	3, 5	
+	−	−	6	
−	+	+	1	
−	+	−	3, 5	70% Co–Ni, 40% Co–Ni
−	−	+	2, 4	
−	−	−	6	Ni

6 types of curves as indicated in Fig. 21, the type depending on the presence or absence of maxima or minima, and on the existence of a crossing of the curve through $\lambda=0$. The situations that give each type of curve are given in Table III, and some materials that fall into various classes are noted.

ACKNOWLEDGMENTS

We are indebted to P. W. Anderson, C. Herring, H. J. Williams, E. A. Nesbitt, and J. K. Galt for valuable discussions on various phases of this work. We are also indebted to S. Millman for valuable comments on the manuscript, and to S. Geschwind and J. F. Dillon, to B. B. Cetlin, J. K. Galt, F. R. Merritt, and W. A. Yager, for permission to quote their values of K_1 obtained from microwave data before publication.

Saturation Hall Constant of Semiconductors

JOHN A. SWANSON
International Business Machines Corporation, Poughkeepsie, New York
(Received January 20, 1955)

The asymptotic value of the Hall constant of semiconductors as the magnetic field becomes very large is shown on general grounds to be $1/(p-n)ec$, where p and n are the concentrations of holes and of electrons respectively; this result is independent of the band shapes in the neighborhood of energy maxima or minima. The unscattered motion of electrons, which determines the saturation Hall constant, is examined by means of the usual crystal momentum force equation, and also through direct investigation of the properties of stationary states in crossed electric and magnetic fields. The saturation magnetoresistance is also discussed, within the framework of the crystal momentum approximation.

INTRODUCTION

THE Hall constant of semiconductors may exhibit considerable variation with magnetic field strength. At vanishingly small fields, the Hall constant of extrinsic semiconductors takes the form

$$R_0=\pm\frac{\mu_H}{\mu}\frac{1}{n_\pm ec}, \tag{1}$$

where $n_\pm$ is the concentration of holes (+) or of electrons (−), μ is the drift mobility, μ_H is the Hall mobility, and e and c have their usual significance. The ratio μ_H/μ in general differs from unity, since electrons in different states may be deflected through different individual Hall angles.

As the magnetic field is increased the Hall constant changes, and may eventually approach a saturation value, R_∞. Theoretical calculations of Hall constant *versus* magnetic field strength have been carried out for certain special types of energy band—those having spherical[1] or ellipsoidal[2] surfaces of constant energy in crystal momentum space. In these instances the saturation Hall constant is

$$R_\infty=\pm 1/n_\pm ec. \tag{2}$$

In this note we shall examine specifically the theory of the saturation Hall constant, and shall attempt to generalize somewhat the conditions under which formula (2) is valid.

I. PHYSICAL INTERPRETATION OF THE SATURATION CONSTANT

At saturation the longitudinal current can be calculated exactly as though the particles were unscattered, that is, as though they followed their natural motion in the electric and magnetic fields.

The reasoning leading to this conclusion, and to the value of the saturation constant itself, is especially straightforward if all the particles have a single effective mass, m^*. The orbits of such particles in crossed electric and magnetic fields are well known; they are cycloids with angular frequency

$$\omega=eH/m^*c \tag{3}$$

and translational velocity

$$v_0=(E/H)c \tag{4}$$

in the direction $(\mathbf{E}\times\mathbf{H})$. A simple proof of this consists in transforming to a coordinate system moving with velocity $\mathbf{v}_0$ relative to the initial system. In this new coordinate system there is no electric field.

Let the mean free time of a particle in a given cycloidal orbit be τ. Then on the average a particle will traverse the angle

$$\theta'=\omega\tau, \tag{5}$$

before being scattered. If θ' is small, the Hall angle θ between current and electric field is given by

$$\theta=\omega\tau_H, \tag{6}$$

where τ_H is a weighted average of the mean free times for different orbits, in general not identical to the average used in calculating the drift mobility. The definition of the Hall mobility,

$$\mu_H=c\lim_{H\to 0}\frac{\theta}{H}, \tag{7}$$

leads to the equation

$$\mu_H=(e/m^*)\tau_H. \tag{8}$$

This abbreviated conventional analysis shows that the Hall constant at low magnetic fields depends on the distribution of mean free times.

When the magnetic field is sufficiently large, θ' for all particles becomes greater than 2π, while θ cannot exceed $\pi/2$. The particles follow their natural motion for one or more revolutions, and hence between collisions possess average velocity v_0. Since θ approaches $\pi/2$, this velocity has the direction of the normal longitudinal current. It is reasonable to suppose that collisions merely carry a particle from one orbit to another without adding any characteristic displacement in the longitudinal direction. In this event the longi-

[1] A. H. Wilson, *Theory of Metals* (Cambridge University Press, Cambridge, 1953), Chap. VIII.
[2] Motoichi Shibuya, Phys. Rev. 95, 1385 (1954).

tudinal current i_l is

$$i_l = \pm n_\pm e v_0, \tag{9}$$

and the saturation Hall constant is clearly given by Eq. (2).

The fact that scattering may be ignored at high fields leads us to make a more general analysis of the motion of electrons in crystals under crossed electric and magnetic fields. Proof that scattering does not affect the saturation Hall constant in the general case will be deferred until Sec. III, since it will be desirable to have the theory of the unscattered motion already in mind when considering this and related questions.

II. MOTION IN CROSSED FIELDS

Orbits in Momentum Space

The particle orbits in momentum space are determined by the force equation

$$d\mathbf{P}/dt = -e[\mathbf{E} + (\mathbf{v}\times\mathbf{H})/c], \tag{10}$$

and the associated velocity equation

$$v_i = \partial W/\partial P_i, \tag{11}$$

where $\mathbf{P}$ is the crystal momentum, $\mathbf{v}$ the velocity, and W the band energy. This description of the motion does not include quantum effects specific to the oscillation of the particle in its orbit. We shall discuss this question in Sec. IV.

In order that $\dot{P}_i$ may be continuous we shall not employ the reduced zone scheme initially, but instead allow P_i to take on all real values. The energy in one band then becomes a periodic function of crystal momentum, i.e., three primitive $\mathbf{P}$-vectors may be defined such that any displacement which is a sum of primitive displacements leaves the energy function unchanged. The images of the orbits in the reduced Brillouin zone will be discussed later.

Let us set up a right-handed coordinate system with orthogonal unit vectors $\mathbf{u}_x$, $\mathbf{u}_y$, and $\mathbf{u}_z$ pointing in the direction of the x, y, and z axes respectively. The directions of the electric and magnetic fields are specified as follows:

$$\mathbf{E} = E\mathbf{u}_z,$$
$$\mathbf{H} = -H\mathbf{u}_z, \tag{12}$$

so that the vector $(\mathbf{E}\times\mathbf{H})$ points in the direction of the y-axis.

The motion is confined to a plane in momentum space

$$P_z = \text{const}. \tag{13}$$

The other component equations of (10),

$$\dot{P}_x = -e[E - v_y H/c], \tag{14}$$

$$\dot{P}_y = -ev_x H/c, \tag{15}$$

can be combined to yield

$$dP_x/dP_y = (cE - v_y H)/v_x H, \tag{16}$$

which in conjunction with Eq. (11) determines the orbit. A more informative equation is obtained by using the well-known relation

$$\dot{W} = \mathbf{v}\cdot d\mathbf{P}/dt, \tag{17}$$

which follows from Eq. (11). Substituting the expression for $d\mathbf{P}/dt$ given by Eq. (10), we have

$$\dot{W} = -eEv_x, \tag{18}$$

and from Eqs. (15) and (18) it follows that

$$dW/dP_y = v_0, \tag{19}$$

where v_0 has the value cE/H, as in Eq. (4). We shall base our further discussion on this relation.

The orbits in P-space, as given by Eqs. (13) and (19), may be visualized as follows: Confining ourselves to the plane $P_z = \text{const}$, let the energy as a function of P_x and P_y be plotted in what was formerly the P_z direction. Then the orbits in P-space are the projections on the plane $P_z = \text{const}$ of the intersection of planes with slope $dW/dP_y = v_0$ with that energy surface.

It is clear that all of the orbits in P-space are either: (a) infinite in extent with average direction along the electric field, or (b) reentrant. For if an orbit were confined to a bounded region of the plane $P_z = \text{const}$ and were not re-entrant, there would have to be an infinite number of distinct intersections of a particular plane of slope (19) with the energy surface, and this is of course impossible in cases of physical interest.

If v_0 is different from zero and not too large, orbits of both type (a) and type (b) exist. The orbits are obviously reentrant near a maximum or minimum of the energy surface (i.e., the surface which represents energies in the $P_z = \text{const}$ plane, and conceived in the preceding as erected above that plane), providing v_0 is small enough. It will be shown later that orbits of type (a) must then also exist. The special case $v_0 = 0$ corresponds to an infinite magnetic field or zero electric field. In this case all orbits are constant energy contours. For $v_0 = \infty$ all the orbits are infinite straight lines, corresponding to motion in a pure electric field.

We are interested primarily in the time average of the velocity of particles in these momentum orbits. Using the angular parentheses to designate the time average of the included quantity, we find from Eqs. (14) and (15) that

$$\langle v_y \rangle = v_0 + \langle \dot{P}_x \rangle c/eH, \tag{20}$$

$$\langle v_x \rangle = -\langle \dot{P}_y \rangle c/eH = 0, \tag{21}$$

the last equation following from the fact that P_y is bounded (since the energy in one zone is bounded). Furthermore, for any re-entrant orbit $\langle \dot{P}_x \rangle$ is zero, and thus

$$\langle v_y \rangle = v_0 \text{ (all reentrant orbits)}. \tag{22}$$

It is Eq. (22) which allows us to realize the more general validity of Eq. (2).

We next consider the images of these orbits in the

reduced Brillouin zone. Let us select one plane P_z=const and one energy plane $W=v_0P_y+$const. Then if there exists no vector in the direction of the electric field which can be expressed as a sum of the primitive momentum displacement vectors, the image of the orbits so defined will be everywhere dense in the fundamental zone or on a surface in the fundamental zone. Let us for simplicity choose a direction for E which contains such a vector. Then the orbits in the topological space of the reduced zone are "closed" curves. We may generate a complete set of such closed orbits by imagining the reduced zone to be completely populated at one instant of time by the array of points representing the stationary states. The orbits subsequently traced out by these points give us our complete set.

Having a complete set of orbits in the reduced zone we can attach a more definite meaning to the terms of the following statement: the sum of the average velocities associated with all orbits is zero. (If several orbits coincide, the contribution of each must, of course, be counted.) This must be so since the total current contributed by a full Brillouin zone is zero. One should recall here that the flow of points in P-space is "incompressible." Since all the reentrant orbits have the average velocity v_0, those orbits which are not reentrant in the infinite P-space have on the average an opposite velocity. Both reentrant and non-re-entrant orbits must exist in general in order that the sum of all average velocities be zero.

Application to the Hall Effect

Let us suppose that the great majority of electrons in a given band have energies less than a certain energy W_0. Consider the volume in extended P-space composed of points associated with energies less than or equal to W_0. If this volume is composed of subvolumes which are finite in extent and disjoint, then the current carried by the corresponding filled orbits in the fundamental zone is $-n_-ev_0$, where n_- is the number of electrons in the band. This follows from the fact that the electrons are confined to these disjoint regions if the velocity v_0 is small enough. If there are several partially filled bands, each satisfying the above criterion, the same formula for the current will apply, with n_- equal to the total number of electrons in such partially filled bands. The image of the disjoint regions on the reduced zone may be either connected or disconnected.

The same result clearly holds for holes. An empty orbit can be considered to carry the current $+ev_0$. If the regions in P-space corresponding to all relevant unfilled states are finite in extent and disjoint, the current is n_+ev_0, where n_+ is the number of unoccupied states.

Experimentally, saturation of the Hall constant is obtainable only if the mean free time is sufficiently large that at available fields most of the particles average at least one traversal of their orbits between scattering collisions. If the electrons are inhomogeneous, i.e., are spread among several bands or different energy minima, the saturation condition may be realized by different groups of electrons at different field strengths. As a result the Hall constant may still change materially after the first apparent approach to saturation.

We conclude this section by pointing out that when both holes and electrons are present the longitudinal current density is

$$i_l=ev_0(n_+-n_-), \tag{23}$$

and the Hall constant is

$$R_\infty=1/(n_+-n_-)ec. \tag{24}$$

This is a well-known formula, now validated more generally. For a near-intrinsic semiconductor the criterion for saturation is somewhat different, as explained in Sec. III.

III. STATISTICAL TREATMENT

In this section, we shall apply the Boltzmann transport equation to the problem of determining the distribution function at high fields. This analysis enables us to show that the high field Hall constant is determined by the free motion of the carriers. The discussion also leads naturally to a study of the saturation magnetoresistance.

Let f be the probability of occupation of states of a band in the reduced zone of P-space. If the concentration of carriers is uniform in space, the Boltzmann equation for this band is

$$\frac{d\mathbf{P}}{dt}\cdot\nabla_P f=-e[\mathbf{E}+(\mathbf{v}\times\mathbf{H})/c]\cdot\nabla_P f=\left[\frac{\partial f}{\partial t}\right]_{\text{coll}}, \tag{25}$$

where the symbol ∇_P denotes the gradient with respect to crystal momentum. The right member of Eq. (25) represents the net rate of change of occupation probability due to scattering collisions.

From Eq. (25) one can deduce in a general manner the behavior of the electric field as the magnetic field increases without limit while the longitudinal current density remains constant. Since the right member of Eq. (25) remains finite while the force $d\mathbf{P}/dt$ becomes infinite, the gradient of the distribution function along the direction of the force must approach zero. Thus the distribution function tends to become constant along the unscattered orbits. It follows that the electric field must become infinite; otherwise, the limiting distribution function would have zero current associated with it. However, the component of the electric field in the direction of $\mathbf{H}$ will remain finite, since a finite electric field is sufficient to cause a current in that direction. In the limit the electric field therefore becomes perpendicular to the magnetic field.

It follows from the above that the current density at

saturation in the plane perpendicular to H is determined by the unscattered motion of the carriers in the magnetic and transverse electric fields. This current, as we have seen, is orthogonal to both fields. Should the magnetic field be applied in a direction not perpendicular to that of the longitudinal current density, a finite electric field in the same direction as the magnetic field is required to produce a balancing current, such that the resultant current lies in the longitudinal direction. The current density due to the unscattered motion of the particles in the magnetic and Hall electric fields is thus the projection on the axis perpendicular to E and to H of the longitudinal current density. It follows that the saturation Hall constant is determined by this unscattered motion, and has the same value for all directions of H, provided that the definition of the Hall constant is generalized in the usual way:

$$\mathbf{E}_H = -R(\mathbf{i}\times\mathbf{H}), \tag{26}$$

in which $\mathbf{E}_H$ is the Hall field, and $\mathbf{i}$ the vector current density.

In the following we shall treat the saturation Hall constant and also the magnetoresistance in a more rigorous fashion for the case in which the magnetic field is perpendicular to the longitudinal current density. We will adopt the same notation and direction for each field as in the previous section.

It is convenient for the analysis to constrain one field to be a constant multiple of the other, i.e.,

$$E = Hv_0/c. \tag{27}$$

The constant v_0 determines a set of unscattered orbits in the crossed fields. Let us define for every such orbit a specific orbit time t' such that t' is the time (measured with respect to an arbitrary fixed reference moment) at which the particle occupies a particular point of its orbit in a magnetic field of unit strength. In a general field H, the corresponding time t is related linearly to t'

$$t' = Ht, \tag{28}$$

and Eq. (25) becomes in terms of t'

$$H df/dt' = [\partial f/\partial t]_{\text{coll}}. \tag{29}$$

In Eq. (29) the right member is a partial derivative with respect to time at a fixed point in momentum space, while the term df/dt' is a total derivative with respect to specific time along the orbit passing through the point in question.

Let us assume that at sufficiently high fields the following expansions are valid:

$$f - f_0 = h_0 + h_1 H^{-1} + h_2 H^{-2} + \cdots, \tag{30}$$

$$[\partial f/\partial t]_{\text{coll}} = s_0 + s_1 H^{-1} + s_2 H^{-2} + \cdots. \tag{31}$$

It will prove useful to define the additional quantities

$$\bar{h}_0 = f_0 + h_0, \tag{32a}$$

$$\bar{h}_i = h_i \quad (i \neq 0). \tag{32b}$$

The quantity $\bar{h}_0$ is the saturation probability distribution function for the particular value of v_0 selected.

Applying Eq. (28) to the expansions (29) and (30) we find that

$$d\bar{h}_{i+1}/dt' = s_i. \tag{33}$$

The first equation of the set is

$$d\bar{h}_0/dt' = 0 \tag{34}$$

showing that $\bar{h}_0$ is constant along each of the orbits determined by v_0. Thus, provided that the series (30) converges, the saturation Hall constant is rigorously shown to be determined by the unscattered motion. It should be noted, however, that Eq. (34) does not determine the distribution of electrons among the orbits. As one would expect, this distribution is determined by the law of scattering.

Let us in fact note that every s_i must satisfy the condition

$$\oint_{v_0} s_i dt' = 0. \tag{35}$$

The loop integral sign indicates that the scattering function s_i is to be integrated around one complete revolution of the particle in one of the orbits determined by the constant v_0. The proof of Eq. (35) consists in integrating Eq. (33) with respect to the time around one of the closed orbits. The integral of the left hand member is the change of h_i around the orbit, which must be zero.

Equation (35) constitutes a condition on h_i, since h_i determines s_i through the scattering law. This condition is needed to specify h_i uniquely, since to any h_i satisfying (33) we can add a function constant along the orbits but otherwise arbitrary, thereby obtaining a new solution. In particular, the value of $\bar{h}_0$ along a given orbit is determined by Eq. (35) as applied to s_0.

Rather than discuss the general case further, we shall suppose that a mean free time exists. The problem is then greatly simplified, since the functions s_i are given in terms of h_i by the relation

$$s_i = -h_i/\tau. \tag{36}$$

The set of equations resulting from substitution of this expression for s_i into Eqs. (33) is easily soluble, and moreover the solution may be found along each orbit independently. For suppose h_i has been determined; then we may determine h_{i+1} by integrating part way around an orbit:

$$h_{i+1}(P) - h_{i+1}(P_0) = -\int_{t_0'}^{t'} h_i dt'/\tau. \tag{37}$$

Here P_0 and t_0' are the coordinates of an arbitrary reference point on the orbit, and P and t' are the coordinates of a general point. Equation (36) determines h_{i+1} along every orbit to within an arbitrary constant. This constant is determined from the con-

dition

$$\oint_{v_0} h_{i+1}dt'/\tau=0, \tag{38}$$

following from Eq. (33).

We next show that the series (30) obtained in this manner converges if H is sufficiently large. We note that the maximum absolute value of h_{i+1} along a particular orbit is less than the maximum absolute value of h_i integrated around the complete orbit. This follows from Eq. (37). Indicating the maximum absolute value along an orbit by the subscript m, we have

$$h_{i+1,m}<h_{i,m}\oint_{v_0} dt'/\tau. \tag{39}$$

Since h_i must take both positive and negative values, the inequality (39) is in general rather strong. It is useful for our purposes to express this inequality in a different form. Taking the line integral with respect to the actual orbit time t we have

$$\frac{h_{i+1,m}}{h_{i,m}H}<\oint_{v_0} dt/\tau. \tag{40}$$

The left member of Eq. (40) is the ratio of the maximum absolute values of adjacent terms in the series (30). Thus the series converges for a given orbit if

$$\oint_{v_0} dt/\tau<1. \tag{41}$$

This condition may be assured for all relevant (occupied) orbits by choosing a sufficiently large magnetic field.

Thus if a mean free time exists, the convergence of series (30) for sufficiently high magnetic fields can be proven rigorously. We note, however, that the series is not convergent for all magnetic fields. In fact, it is not difficult to show that for the condition

$$\oint_{v_0} dt/\tau>2\pi \tag{42}$$

the series diverges.

When the left member of inequality (41) is very small compared to unity, it is approximately equal to the probability of the occurrence of a scattering collision during the time of one revolution of the particle. We may consider the maximum absolute value (along an orbit) of the ith term in the series (30) as a rough measure of the ith order current contributed by a particle in that orbit. Applying inequality (40) to the first and second terms of the series (30), we see that the major part of the current contributed by particles occupying a given band and region of momentum space is due to the unscattered motion of these particles if the probability per revolution of scattering in that band and region of momentum space is small.

It would be wrong, however, to conclude that the saturation Hall constant is effectively reached when the preceding sentence is true for all relevant bands and regions of momentum space. The condition for the saturation of the Hall constant is that the current density due to the first term of (30) is much larger than the current density due to the second term. But if both nearly empty and nearly filled bands must be considered, the contribution to the current of the holes balances the contribution of an equal number of electrons as far as the current due to the unscattered motion is concerned. The magnitude of the current due to h_0 is therefore reduced, and the current associated with h_1 may continue to be dominant or appreciable at fields much higher than those required to reduce the probability of scattering per revolution to a very small quantity for all orbits. This enables us to understand the implausible discontinuity exhibited by formula (24) at the intrinsic point. At any given finite field the Hall constant is continuous as a function of the net impurity concentration, but in a region of concentrations close to intrinsic it will not have reached its saturation value. Only in the limit of infinite magnetic fields will the Hall constant have reached saturation for all impurity contents, and therefore it is only in this unrealizable case that the discontinuity exists.

To conclude this section we discuss briefly the possibility of expanding the distribution function in the following manner:

$$f=f_0+E(g_1H^{-1}+g_2H^{-2}+\cdots). \tag{43}$$

This expansion is incomplete in the sense that terms of order higher than the first in the electric field are omitted; nevertheless the differential equation governing the g's can be solved in a manner similar to that employed above, with the integrals in this case taken along constant energy contours. The previous expansion is complete since it contains all powers of E by virtue of the constraint $Ec=v_0H$. This constraint does not detract from the generality of the solution, since v_0 is arbitrary.

Saturation Magnetoresistance

Let the Hall angle θ be expressed as

$$\theta=\pi/2-\phi. \tag{44}$$

Then near the saturation of the Hall constant

$$\phi\doteq i_E/i_l, \tag{45}$$

where i_l is the current due to the free motion of the particles, and i_E is the current in the direction of $\mathbf{E}$. The electric field E_l in the direction of the longitudinal current can be expressed in terms of ϕ:

$$E_l\doteq E\phi. \tag{46}$$

The resistivity at infinite magnetic field, ρ_∞, which

we shall call the "saturation resistivity," is given by

$$\rho_\infty = E_t/i_1 = \phi E/i_1 = R_\infty H i_E/i_1. \quad (47)$$

In general the difference in resistivities

$$\Delta\rho = \rho_\infty - \rho_0,$$

where ρ_0 is the resistivity at zero magnetic field, does not vanish, and a "saturation magnetoresistance" exists, proportional to this difference. Since h_0 does not give rise to current in the direction of the electric field, the saturation magnetoresistance is determined by h_1.

The form of the functions $h_0(P)$ and $h_1(P)$ depends on the value of v_0. Thus it may be that the saturation magnetoresistance is current-dependent. In the following we shall be concerned with the limiting value of the magnetoresistance as the longitudinal current density is reduced to zero. The saturation magnetoresistance should not deviate appreciably from this limiting value if the longitudinal drift velocity is much less than the thermal velocity associated with each energy minimum.

We wish to obtain a formula for h_1, but in order to do so it is necessary first to determine h_0. Let us concentrate our attention upon a particular orbit and a particular constant energy contour (energy W_0) which intersects it at a point P_0. Then, if v_0 is sufficiently small,

$$f_0(W) \doteq f_0(W_0) + (\partial f_0/\partial W)(W - W_0), \quad (48)$$

where W is the energy along the orbit. Now consider a point on the energy contour and the adjacent point on the orbit which has the same P_x coordinate. Then, in view of (19),

$$f_0(W) - f_0(W_0) = (\partial f_0/\partial W) v_0 (P_y - P_{0y}). \quad (49)$$

Since P_{0y} is arbitrary we may set $h_0(W)$ equal to $f_0(W_0)$. Then Eq. (37) (for $i = -1$) constitutes a condition on P_{0y}:

$$\oint_{v_0} (P_y - P_{0y}) dt'/\tau = 0. \quad (50)$$

Since we are specializing to cases in which v_0 is small, the condition (50) may be approximated by

$$\oint_W (P_y - P_{0y}) dt'/\tau = 0, \quad (51)$$

an integral around a constant energy contour. Rewriting Eq. (49) we have

$$h_0(W) = f_0(W) - (\partial f_0/\partial W) v_0 (P_y - P_{0y}). \quad (52)$$

The two terms on the right-hand side of this equation must correspond to the first two terms of the expansion (43). In fact, had we solved the latter expansion we should have found that

$$g_1 = -(\partial f_0/\partial W)(P_y - P_{0y}), \quad (53)$$

where P_{0y} is determined by Eq. (51). One may employ (52) as an adequate approximation of the correct function near each energy minimum when v_0 is small.

The resultant formula for h_1 is

$$h_1(P) - h_1(P_0) = \frac{\partial f_0}{\partial W} v_0 \int_{t_0'}^{t'} (P_y - P_{0y}) dt'/\tau. \quad (54)$$

To Eq. (54) must be added the usual condition

$$\oint_W h_1 dt' = 0. \quad (55)$$

Equations (50), (53), and (54) determine the saturation magnetoresistance. Should the contours of constant energy in a given plane $P_x = \text{const}$ have a center of inversion in the point of maximum or minimum energy, and should the mean free time possess the same symmetry properties, the problem is further simplified. For then the constant P_{0y} is the same for all the contours in the plane $P_x = \text{const}$, and in fact is equal to the P_y-coordinate of the center of inversion.[8]

It should be noted that this method of finding the magnetoresistance is related to Shockley's "tube integral" method[4] discussed further in the Appendix.

The quantity

$$g_2 = h_1 c/v_0 \quad (56)$$

is independent of v_0. Let I be the operation (consisting of the usual integration over P-space) which, acting on a distribution function, yields the value of the current density in the x-direction (direction of $\mathbf{E}$). Then it is easy to see by virtue of Eq. (47) that

$$\rho_\infty = R_\infty^2 I g_2. \quad (57)$$

IV. STATIONARY STATES IN CROSSED FIELDS

Thus far we have described the unscattered motion of an electron in a periodic potential under crossed electric and magnetic fields in terms of the crystal momentum, i.e., in terms of the stationary states which exist when these fields are absent. In this approximation the electric and magnetic fields are treated as perturbations, a procedure which loses its justification when these fields become excessively large. It is then more appropriate to discuss the motion of the electron in terms of the true stationary states which exist in the presence of the crossed fields. We shall inyestigate the properties of such stationary states which are relevant to the saturation Hall constant, but not those relevant specifically to the saturation magnetoresistance. That is, we shall attempt to calculate the expectation value

[8] With constant P_{0y}, formula (53) represents a displacement of the equilibrium distribution in P-space, and can be obtained by simpler arguments if one assumes that the effect of electric and magnetic fields is to produce such a displacement. See the paper of M. Kohler, in which the saturation magnetoresistance of metals is regarded from a point of view similar to ours [M. Kohler, Ann. Physik **38**, 283 (1940)].

[4] W. Shockley, Phys. Rev. **79**, 191 (1950).

of velocity of an electron in such a state, but not the probabilities of scattering into other states.

In the following we shall find it convenient to discriminate between "restricted band states" and general states. The term "restricted band state" shall refer to states resulting from a superposition of states primarily from one band or group of degenerate bands. Though not, in general, stationary, such states may be "quasi-stationary" in the sense that the probability density is essentially constant over ordinary time periods. Since the probability of field-induced transitions to higher bands is usually small, most electrons will in fact occupy such states. From the familiar result that a full band carries no current it follows that not all restricted band states can have the same expectation value of velocity. However, this conclusion does not apply to the true stationary states.

Before beginning a direct calculation of the velocity expectation value we take cognizance of an exact solution previously obtained for a special case. The problem of charged particles in crossed electric and magnetic fields in free space has already been treated quantum mechanically by Titeica,[5] whose work was concerned with the magnetoresistance of metals. He found that stationary states exist having the form

$$\psi(x,y,z)=f_n(x-x_0)e^{(i/\hbar)m(v_0-\omega x_0)y}e^{ikz\omega}, \tag{58}$$

in which $f_n(x)$ is the harmonic oscillator wave function corresponding to the angular frequency $\omega=eH/mc$ and quantum number n. The average position of the electron, x_0, is arbitrary. It is easily seen by direct application of the velocity operator that the average velocity is v_0.

In this case we see that each wave function belongs to a class, the members of which differ only by relative translations along the x-axis and a corresponding phase factor. Such classes exist in general, a fact which enables us to deduce the expectation value of velocity for stationary states when the electron is contained in a crystal, as shown below.

The Hamiltonian for our problem is

$$H=\frac{1}{2m}\left[p_x^2+\left(p_y-\frac{eH}{c}x\right)^2+p_z^2\right]+eEx+V_p(x,y,z), \tag{59}$$

where V_p is the periodic potential. We have chosen the gauge which gives to the Hamiltonian the maximum translational symmetry in the y-direction. We note in passing that if the y- and z-directions are collinear with lattice displacement vectors the solution of Schrödinger's equation can have a form very similar to (58):

$$\psi(x,y,z)=\phi_n(x,y,z)e^{ik_y y}e^{ikz\omega}, \tag{60}$$

where ϕ_n is periodic in the y- and z-directions.

It will now be shown that, given any solution of Eq. (59) with expectation value of the x-coordinate $\langle x\rangle=x_0$ and energy W_0, infinitely many solutions exist of identical form except for a phase factor, having each a value of $\langle x\rangle$ and a related energy $W_0+\delta W$ such that

$$eE(\langle x\rangle-x_0)=\delta W. \tag{61}$$

Let $\mathbf{a}$ be a general lattice displacement vector. Then, simultaneously with the coordinate transformation

$$\mathbf{r}=\mathbf{r}'+\mathbf{a}, \tag{62}$$

let us introduce the gauge transformation

$$A_y=A_y'-(eH/c)a_x. \tag{63}$$

In the new coordinate system and gauge the original Hamiltonian is recovered except for the additive term eEa_x. Thus if $\psi(\mathbf{r})$ is a solution in the old gauge, $\psi(\mathbf{r}')$ is a solution in the new gauge.

It follows from gauge transformation theory and may easily be verified directly, that the new wave functions in the original coordinate system and gauge have the form

$$\psi(\mathbf{r})=\psi(\mathbf{r}-\mathbf{a})e^{(i/\hbar)\delta P_y y}, \tag{64}$$

where $\delta P_y=eHa_x/c$.

Let us suppose that a wave packet has been constructed by superposition of states of the form (64). Then the group velocity of this packet is

$$v_y=\delta W/\delta P_y=v_0, \tag{65}$$

which is consequently the expectation value of velocity for the component states.

This result has been shown to apply to all true stationary states. We will show below that it applies also to all quasi-stationary restricted band states satisfying a certain condition. This condition is that the restricted band states in question can be generated from stationary states in the magnetic field alone by regarding the electric field as a perturbation. It is clear that the unperturbed states must be localized with respect to the x-coordinate and must have essentially uniform probability density in the y-direction in order that the electric field constitute a perturbation; but it is probable that in general a complete set of states in the magnetic field can be found having this property. The unperturbed states are automatically restricted band states. Consequently, the perturbed states, if they exist, are also restricted band states.

The expectation value of velocity for the perturbed states will now be calculated directly. Letting the symbol δ indicate the change in the conjoined quantity induced by the perturbation, we have

$$\delta\langle v_y\rangle=\frac{1}{m}\left(\delta\langle p_y\rangle-\frac{eH}{c}\delta\langle x\rangle\right). \tag{66}$$

But to first order in electric field we have

$$\delta\langle p_y\rangle=\sum_n{}'\frac{H_{0n}'p_{y n0}+p_{y n0}H_{n0}'}{W_0-W_n}, \tag{67}$$

[5] S. Titeica, Ann. Physik 22, 129 (1935).

and

$$\delta\langle x\rangle=\sum_n{}'\frac{H_{0n}''x_{n0}+x_{0n}H_{n0}''}{W_0-W_n}, \tag{68}$$

where W_n is the energy of the nth unperturbed state and $H_{0n}'=H_{0n}''$ are the matrix elements of the perturbation. The sum is taken over all values of n except $n=0$, this value indicating the initial unperturbed state, $\psi_0(x,y,z)$.

The perturbation H' is eEx. However, on transforming to a new coordinate system defined by

$$x=x'-mc^2E/eH^2, \tag{69a}$$

$$\mathbf{r}=\mathbf{r}'+\mathbf{a}, \tag{69b}$$

where $\mathbf{a}$ is a lattice displacement vector, the term eEx drops out of the Hamiltonian (59), and is replaced by the term v_0p_y. In this new coordinate system one can therefore treat v_0p_y as the perturbation. Out of all the states having the same form except for a phase factor and a displacement let us choose as the unperturbed state in this new coordinate system $\psi_0(x',y',z')$. Then the same perturbed state will be obtained in both cases. This is easily seen if ψ_0 has the form (60). For neither perturbation changes this form nor the value of k_y. But for any given orientation of the lattice relative to the magnetic field we may choose another orientation arbitrarily close to the first such that the wave functions may take the form (60) in the new orientation. [It should also be noted that a succeeding infinitesimal rotation about the y-axis insures that the discrete set of values of E allowed by Eqs. (69) is everywhere dense. The restriction to a discrete set is therefore not significant.]

Now if we compute $\delta\langle p_y\rangle$ in the original coordinate system ($H'=eEx$) and $\delta\langle x\rangle$ in the second coordinate system ($H''=v_0p_y$), the resultant expressions when substituted into Eq. (66) cancel. However, we must correct $\delta\langle x\rangle$ by the amount of the coordinate shift $-mc^2E/eH^2$, and this gives us finally the desired result

$$\delta\langle v_y\rangle=v_0. \tag{70}$$

Relation to Orbits in Momentum Space

Let us consider the classical ensemble of particles satisfying the equations of motion (10) and (11), all occupying a given orbit in crystal momentum space and all having the same average value of x-position, but having random y- and z-coordinates. The probability of finding a particle at a given location is stationary for such an ensemble. We shall say that a quasi-stationary restricted band state "corresponds" to an orbit in momentum space if the expectation values of energy and of position are the same for the wave function as for an ensemble of particles in the given orbit. This notion of correspondence allows us to connect the results of the present section with those of Part II.

Not every orbit will possess a corresponding wave function. The localization of the electron by the magnetic field, which leads to the displacement property expressed by Eq. (64), leads in an intimately related way also to the discrete quantization of the energy of an electron having a given expectation value of position. Thus only orbits belonging to a discrete set possess corresponding wave functions.

This quantum effect, regarded from the viewpoint of the crystal momentum approximation, may be called "orbital quantization." The orbital motion of the electron in space results in quantum mechanical interferences which act to select those particular "orbits" for which the interferences are constructive. A first approximation to the resultant energy levels may be found by regarding P_x and x as conjugate oscillatory variables to which the Bohr-Wilson-Sommerfeld phase integral condition may be applied. Through use of Eq. (15) this condition is easily transformed to

$$-\oint P_x dP_x=\oint P_x dP_y=\frac{ch}{eH}(n+\tfrac{1}{2}). \tag{71}$$

The integrals are equal to the area contained by the orbit. It is interesting to note that this condition yields the correct energy levels for the free particle and for the "pseudoparticle" with different longitudinal and transverse masses.

In cases in which bands overlap, the unperturbed wave functions, as stated earlier, must be assumed to involve superpositions of states from each of the degenerate bands. If there should be degeneracy among the unperturbed states, the correct zeroth-order wave functions are to be selected for use in the perturbation calculation. The resultant quasi-stationary states correspond simultaneously to several orbits of equal energy in different bands.

It is reasonable to suppose that as the electric field approaches zero, perturbation theory should become applicable to almost all quasi-stationary states. However, we should like to know to what states in particular the theory is applicable at given values of field strength. We shall not attempt to answer this question rigorously, but instead shall make use of a plausible argument suggested by the correspondence developed here. It is natural to suppose that if the motion of the electron in a magnetic field, as it is described by means of the crystal momentum approximation, is merely perturbed by the electric field, the actual motion is correspondingly perturbed by the electric field, and not altered essentially. But the reentrant orbits are perturbed constant energy contours, while the nonreentrant orbits are altered in essential form. Thus we should expect reentrant orbits to correspond to perturbed states, and nonreentrant orbits to correspond to those "extraordinary" restricted band states which cannot be generated by means of perturbation theory from states stationary in the magnetic field alone.

Such extraordinary states, since they do not occur near the top or bottom of a band, are not likely to be of interest in connection with the semiconductor Hall effect. However, they do play an interesting role in the one-electron theory of the full band. By setting the expectation value of acceleration equal to zero for a quasi-stationary state, it is easily seen that the velocity expectation value is v_0 if and only if the average force exerted by the lattice on the electron is zero; the corresponding statement in the crystal momentum approximation is contained in Eq. (20). In a full band the electrons must, however, exert a force on the lattice ions equal and opposite to the direct force due to the electric field. This force is exerted by electrons in the extraordinary states. Correspondingly, the condition that formula (24) hold is that the relevant electrons and holes exert no average force on the lattice.

In conclusion we point out that at sufficiently high magnetic fields, given constant longitudinal current, a significant number of electrons will make transitions to higher bands. This phenomenon, which is essentially Zener breakdown due to the transverse electric field, could in principle be observed, since it would not be preceded by avalanche breakdown.

APPENDIX. INTEGRAL SOLUTIONS OF THE BOLTZMANN EQUATION

If a mean free time exists, the solution of Eq. (25) can be written in integral form for arbitrary values of H. We have used a series expansion in the preceding work since it simplifies the discussion of saturation effects. The integral solution is given here for the sake of completeness.

The basic idea of this solution is due to Shockley,[4] though his treatment is somewhat generalized here. Let the magnetic field, $\mathbf{H}$, have an arbitrary magnitude, and an arbitrary orientation with respect to the longitudinal current density. The electric field, $\mathbf{E}$, will have a component $\mathbf{E}_1$ parallel to $\mathbf{H}$, and a component $\mathbf{E}_2$ perpendicular to $\mathbf{H}$. The set of orbits in momentum space of which we shall speak are those defined by $\mathbf{E}_2$ and $\mathbf{H}$. Generalizing notation employed by Shockley in treating constant energy contours, we define an angle variable θ for each orbit such that $\theta=2\pi t/T$, where t is the time corresponding to a point of the orbit, and T is the period of one revolution. We further define

$$\omega=2\pi/T,$$
$$g=f-f_0.$$

Then Eq. (25) can be written in the form:

$$\frac{dg}{d\theta}+\frac{g}{\omega\tau}=\frac{\partial f_0}{\partial W}\frac{e}{\omega}\mathbf{E}\cdot\mathbf{v},$$

in which a term $e\mathbf{E}_1\cdot\nabla_P g$ has been dropped, since it is small compared to $e\mathbf{E}_1\cdot\nabla_P f_0$, a quantity included in the right member. This linear equation must be solved subject to the boundary condition $g(\theta)=g(\theta+2\pi)$. Such a solution is given by the expression

$$g(\theta)=A\int_{\theta-2\pi}^{\theta}f_0(1-f_0)\mathbf{E}\cdot\mathbf{v}(\theta')\exp\left[-\int_{\theta'}^{\theta}d\theta''/\omega\tau\right]d\theta',$$

where

$$A=-e(kT\omega)^{-1}\left[1-\exp\left(-\oint d\theta''/\omega\tau\right)\right]^{-1},$$

all integrals being taken along the orbit. The equivalent quantity $-f_0(1-f_0)/kT$ has been substituted for $\partial f_0/\partial W$. Shockley obtained a solution correct to first order in the transverse electric field by integrating instead along constant energy contours. This approximation should be adequate for normal longitudinal drift velocities. In this case θ is the angle for zero electric field, and the factor $f_0(1-f_0)$ may be placed outside the integral sign. One may integrate $g(\theta)v$ around the contour in order to obtain the contribution to the current of states along the contour. The resultant multiple integral, though of somewhat different form, is equivalent to Shockley's integral.

By expanding the exponentials in power series one effectively achieves an expansion in terms of $1/H$. However, the simple results regarding saturation quantities obtained by the direct power series expansion in conjunction with a discussion of the orbits are not apparent from these integrals.

Thermomagnetic Anisotropy

JEROME I. KAPLAN
United States Naval Research Laboratory, Washington, D. C.
(Received April 20, 1955)

A solution of the Boltzmann transport equation in a solid is given for applied electric and magnetic fields and a temperature gradient. The energy surfaces are assumed to be ellipsoids and the density of states is such that Fermi statistics are applicable. It is shown that the calculated kinetic reciprocal coefficients which define the electric and heat currents satisfy the generalized Onsager reciprocal relations.

INTRODUCTION

A METHOD is given for the solution of the Boltzmann transport equation in a solid with applied electric and magnetic fields and a temperature gradient. The assumptions made are that Fermi statistics are applicable, all quantities of interest in the problem are slowly varying with respect to the energy in the region of the Fermi surface, the energy surfaces in the crystal can be interpreted as ellipsoids, a single energy-independent parameter defines the relaxation time for each ellipsoid, there are no coupling terms between ellipsoids, and Ohm's law is obeyed.

The procedure is to solve the Boltzmann transport equation for each ellipsoid in a coordinate system in which the ellipsoid is diagonal. The calculated electric current **J** and heat current **w** are then transformed back to the laboratory coordinates. The assumed independence of each ellipsoid means that the contribution from each ellipsoid in the laboratory system will be additive.

PROCEDURE

The Boltzmann transport equation as given in Wilson[1] has the form

$$-(e/\hbar)[\mathbf{E}\times(1/c)(\mathbf{v}\times\mathbf{H})]\cdot\nabla_{\mathbf{k}}f+\mathbf{v}\cdot\nabla_{\mathbf{r}}f=-(f-f_0)/\tau, \quad (1)$$

where **E** and **H** are the electric and magnetic fields respectively, f_0 the Fermi distribution function with no applied fields, f the perturbed distribution function, and τ the relaxation time. The other constants are taken to have their usual meaning.

Let $f=f_0+\chi\partial f_0/\partial\epsilon$, where ϵ is energy. Then Eq. (1) becomes, to first order in the perturbing fields,

$$-\frac{e}{\hbar}(\mathbf{E}\cdot\nabla_{\mathbf{k}}\epsilon)+\frac{e}{\hbar c}\mathbf{H}\cdot(\nabla_{\mathbf{k}}\epsilon\times\nabla_{\mathbf{k}}\chi)+\frac{1}{\hbar}(\nabla_{\mathbf{k}}\epsilon\cdot\nabla_{\mathbf{r}}T)\left[T\frac{\partial}{\partial T}\left(\frac{\epsilon-\zeta}{T}\right)\right]=-\frac{\chi}{\tau}, \quad (2)$$

where T is the temperature and ζ the Fermi energy. Assuming the applied fields are arbitrary, Eq. (2) expanded becomes

$$\begin{aligned}&-\frac{e}{\hbar}\left(E_x\frac{\partial\epsilon}{\partial k_x}+E_y\frac{\partial\epsilon}{\partial k_y}+E_z\frac{\partial\epsilon}{\partial k_z}\right)\\&+\frac{e}{\hbar c}\left\{H_x\left(\frac{\partial\epsilon}{\partial k_y}\frac{\partial\chi}{\partial k_z}-\frac{\partial\epsilon}{\partial k_z}\frac{\partial\chi}{\partial k_y}\right)\right.\\&+H_y\left(\frac{\partial\epsilon}{\partial k_z}\frac{\partial\chi}{\partial k_x}-\frac{\partial\epsilon}{\partial k_x}\frac{\partial\chi}{\partial k_z}\right)\\&\left.+H_z\left(\frac{\partial\epsilon}{\partial k_x}\frac{\partial\chi}{\partial k_y}-\frac{\partial\epsilon}{\partial k_y}\frac{\partial\chi}{\partial k_x}\right)\right\}\\&+\frac{T}{\hbar}\frac{\partial}{\partial T}\left(\frac{\epsilon-\zeta}{T}\right)\left(\frac{\partial\epsilon}{\partial k_x}\frac{\partial T}{\partial x}+\frac{\partial\epsilon}{\partial k_y}\frac{\partial T}{\partial y}+\frac{\partial\epsilon}{\partial k_z}\frac{\partial T}{\partial z}\right)\\&=-\frac{\chi}{\tau}. \quad (3)\end{aligned}$$

The assumed diagonalized energy surface is written as

$$\epsilon=a_1k_x^2+a_2k_y^2+a_3k_z^2, \quad (4)$$

and χ is taken to be

$$\chi=Ak_x+Bk_y+Ck_z. \quad (5)$$

Substitution of both of these expressions into Eq. (3) gives the result

$$\begin{aligned}&k_x[A-2a_1C\mathfrak{H}_y+2a_1B\mathfrak{H}_z-2a_1\mathfrak{F}_x+2a_1T_x]\\&+k_y[B+2a_2C\mathfrak{H}_x-2a_2A\mathfrak{H}_z-2a_2\mathfrak{F}_y+2a_2T_y]\\&+k_z[C-2a_3B\mathfrak{H}_x+2a_3A\mathfrak{H}_y-2a_3\mathfrak{F}_z+2a_3T_z]=0, \quad (6)\end{aligned}$$

where

$$\mathfrak{F}_i=(\tau e/\hbar)E_i, \quad \mathfrak{H}_i=(\tau e/\hbar c)H_i,$$

$$T_i=\frac{\tau T}{\hbar}\frac{\partial}{\partial T}\left(\frac{\epsilon-\zeta}{T}\right)\frac{\partial T}{\partial x_i}.$$

To satisfy Eq. (6) for a general **k**, the coefficients of k_i must each be equal to zero. The resulting set of three simultaneous equations are solved for A, B, and C with

[1] A. H. Wilson, *Theory of Metals* (Cambridge University Press, London, 1953), Eq. (8·11·12).

the result that

$$AS=(\mathfrak{F}_x-T_x)\alpha_{11}+(\mathfrak{F}_y-T_y)\alpha_{12}+(\mathfrak{F}_z-T_z)\alpha_{13},$$
$$BS=(\mathfrak{F}_x-T_x)\alpha_{21}+(\mathfrak{F}_y-T_y)\alpha_{22}+(\mathfrak{F}_z-T_z)\alpha_{23}, \quad (7)$$
$$CS=(\mathfrak{F}_x-T_x)\alpha_{31}+(\mathfrak{F}_y-T_y)\alpha_{32}+(\mathfrak{F}_z-T_z)\alpha_{33},$$

where

$$S=1+4a_2a_3\mathfrak{H}_x^2+4a_1a_3\mathfrak{H}_y^2+4a_1a_2\mathfrak{H}_z^2,$$
$$\alpha_{ii}=2a_i(1+4a_ja_k\mathfrak{H}_i^2),$$
$$\alpha_{ij}=2a_ia_j(2\mathfrak{H}_k\epsilon_{ijk}+4a_k\mathfrak{H}_i\mathfrak{H}_j).$$

ϵ_{ijk} is defined as 1 for even permutations of ijk and -1 for odd permutations. The expressions for the electric and heat current,

$$\mathbf{J}=\frac{e}{4\pi^3\hbar}\int\nabla_k\epsilon f d\mathbf{k}, \quad \mathbf{w}=\frac{1}{4\pi^3\hbar}\int\nabla_k\epsilon\epsilon f d\mathbf{k}, \quad (8)$$

become, in component form,

$$J_x=\frac{e}{4\pi^3\hbar}\int 2a_1Ak_x^2\frac{\partial f_0}{\partial\epsilon}d\mathbf{k},$$
$$J_y=\frac{e}{4\pi^3\hbar}\int 2a_2Bk_y^2\frac{\partial f_0}{\partial\epsilon}d\mathbf{k}, \quad (9a)$$
$$J_z=\frac{e}{4\pi^3\hbar}\int 2a_3Ck_z^2\frac{\partial f_0}{\partial\epsilon}d\mathbf{k},$$
$$w_x=\frac{1}{4\pi^3\hbar}\int 2a_1A\epsilon k_x^2\frac{\partial f_0}{\partial\epsilon}d\mathbf{k},$$
$$w_y=\frac{1}{4\pi^3\hbar}\int 2a_2B\epsilon k_y^2\frac{\partial f_0}{\partial\epsilon}d\mathbf{k}, \quad (9b)$$
$$w_z=\frac{1}{4\pi^3\hbar}\int 2a_3C\epsilon k_z^2\frac{\partial f_0}{\partial\epsilon}d\mathbf{k},$$

where the cross terms k_ik_j have been dropped as they average to zero.

The integration of Eq. (9) is simplified by making a change of coordinates so that in the new coordinates the energy surface will be spherical. Let $(\hbar^2/2m)k_i'^2=a_ik_i^2$ and then Eq. (9) will be the same except for a factor $(\hbar/2m)^{\frac{1}{2}}a_i^{-1}(a_1a_2a_3)^{-\frac{1}{2}}$ out in front of each current component. The advantage of the change of variable is that $(\hbar^2/2m)k_i'^2$ can now be replaced for purposes of integration by $\epsilon/3$. The resulting integrals will then be over the single variable ϵ. Defining

$$\frac{S^{-1}(\hbar^2/2m)^{\frac{1}{2}}T}{6\pi^2\hbar(a_1a_2a_3)^{\frac{1}{2}}}\int\epsilon^n\frac{\partial f_0}{\partial\epsilon}d\mathbf{k}'=\frac{4\pi TS^{-1}}{6\pi^2\hbar(a_1a_2a_3)^{\frac{1}{2}}}\int\epsilon^{n+1}\frac{\partial f_0}{\partial\epsilon}d\epsilon \quad (10)$$

as K_n,[2] Eq. (9) becomes

$$J_i=\sum_j[eK_1\alpha_{ij}\nabla_j\mu-eK_2\alpha_{ij}\nabla_j(1/T)]$$
$$w_i=\sum_j\left[\frac{K_2\alpha_{ij}}{T}\nabla_j\mu-K_2\alpha_{ij}\nabla_j(1/T)\right], \quad (11)$$

where

$$\nabla_j\mu=\mathfrak{F}_j+\frac{\tau T}{\hbar}\frac{\partial}{\partial T}(\zeta/T)\frac{\partial T}{\partial x_j}, \quad \nabla_j(1/T)=-\frac{\tau}{T^2\hbar}\frac{\partial T}{\partial x_j}.$$

Equations (11) have the form which Kohler[3] has shown the electric and heat currents must have when calculated from the Boltzmann transport equation if one assumes that the transition rate between two states of an electron is microscopically reversible and the same with the field on as at equilibrium. Based on quite similar arguments, it has been shown that Eqs. (11) are a general consequence of the Onsager[4] reciprocal relations generalized to anisotropic[5] media.

For materials where there are multiple ellipsoidal energy surfaces, such as those postulated to explain the de Haas-van Alphen[6] effect, Eqs. (11) must be modified to account for the fact that the ellipsoids in general will not be diagonal in the laboratory coordinate system. Therefore, there will be a set of Eqs. (11) for each ellipsoid which must be transformed to the laboratory system. In matrix form, the transformation can be written as

$$J_i'=\sum_{\mu,j}R_{ij}^{\mu}J_j^{\mu}(\mathfrak{F},\mathfrak{H},\nabla T)$$
$$=\sum_{\mu,j}R_{ij}^{\mu}J_j^{\mu}(R^{-1}\mathfrak{F}',R^{-1}\mathfrak{H}',R^{-1}\nabla T')$$
$$w_i=\sum_{\mu,j}R_{ij}^{\mu}w_j^{\mu}(\mathfrak{F},\mathfrak{H},\nabla T)$$
$$=\sum_{\mu,j}R_{ij}^{\mu}w_j^{\mu}(R^{-1}\mathfrak{F}',R^{-1}\mathfrak{H}',R^{-1}\nabla T'),$$

where the summation μ is over the number of ellipsoids, the R^{μ} is the transformation matrix for the μth ellipsoid.

In general, there will be twenty-one independent kinetic coefficients defining the three electric and three heat currents components. Callen[7] has shown that in a planar isotropic system where there are ten independent kinetic coefficients, the many thermomagnetic effects can simply be described in terms of these coefficients. The generalization to the full set of coefficients is straightforward. Wilson[8] shows how the results for the

[2] See reference 1, Eq. (1·7·2).
[3] M. Kohler, Ann. Physik **40**, 601 (1941).
[4] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).
[5] Fieschi, DeGroot, and Pazur, Physica **20**, 67 (1954).
[6] D. Shoenberg, Proc. Roy. Soc. (London) **A170**, 341 (1939).
[7] H. Callen, Phys. Rev. **85**, 16 (1952).
[8] See reference 1, p. 211.

combination of holes and electrons can be calculated by a simple modification of the above procedure.

CONCLUSION

A method is given for the calculation of all the thermomagnetic effects based on a multiple-ellipsoidal energy surface model. It is shown that the kinetic reciprocal coefficients satisfy the generalized Onsager[4,5] reciprocal relations. It is thought that the formulation might be used to correlate the thermomagnetic properties of a metal with its effective masses as found from de Haas-van Alphen effects.[6]

ACKNOWLEDGMENTS

I wish to thank Dr. Martin Steele for helpful discussions.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Electrical Properties of Germanium-Silicon Alloys*

ALFRED LEVITAS†
Sylvania Electric Products, Inc., Ipswich, Massachusetts
(Received May 25, 1955)

Resistivity measurements have been performed from 300° to 800° Kelvin on homogeneous germanium-silicon alloys. The intrinsic lines for these alloys have been established. Hall and resistivity measurements have been performed on p-type germanium-silicon alloys from 77° to 300°K. The variation of mobility with composition of the alloy is indicative of alloy scattering.

1. INTRODUCTION

THE preparation of homogeneous alloys of germanium and silicon was reported originally by Stohr and Klemm[1] in 1939. More recently, these alloys have been prepared by Wang and Alexander[2] and by Johnson and Christian.[3]

The availability of these alloys has stimulated considerable interest in their properties. The alloys are substitutional and the diamond crystal structure is maintained. Nevertheless, the periodicity of the lattice is disturbed by the random arrangement of the constituent atoms on the lattice sites. For this reason certain properties of germanium alone and silicon alone need to be re-examined for the alloys.

The usual energy band concept, based on a one-electron wave equation approximation, depends strongly on the assumption of a periodic potential in the crystal. In the alloys, with periodicity disturbed, one might expect that a distinct band structure would not obtain. The optical determination of forbidden band width of the alloys by Johnson and Christian[3] are based on absorption curves with slopes which are similar to that for germanium alone. This seems to imply that, at least so far as the forbidden band width is concerned, a distinct band structure does exist for each alloy. If a distinct band structure does indeed exist for the alloys, it is then possible to speculate, as Herman[4] has, on the manner in which the band structure of germanium deforms to become the band structure of silicon as the composition of the alloy is changed from 0% silicon to 100% silicon.

On the basis of Hall mobility data which will be described in the present work, it seems likely that the aperiodicity of the lattice manifests itself on a scale which strongly affects the mobilities of carriers in the alloys. Brooks[5] has suggested that in a random solid solution the composition will vary from region to region by virtue merely of normal statistical fluctuation. This will produce a hill and valley structure in the band edges and deformations in the energy bands which can be treated in a manner entirely analogous to the deformation potentials which arise in the theory of thermal scattering. If the mean free path for scattering by this mechanism is calculated, one loses a factor $1/T$ which appears in the calculation of the mean free path in thermal scattering. Thus the alloy scattering is treated as a "frozen-in" thermal scattering, and the anticipated temperature dependance for this scattering mechanism is $T^{-0.5}$ in contrast to the dependance $T^{-1.5}$ expected for thermal scattering. An important parameter determining the magnitude of the alloy scattering is the rate of shift of the band edge with composition of the alloy. This parameter is analogous to the parameter describing the rate of shift of the band edge with deformation in the theory of thermal scattering. Brooks

* Supported in part by Contract with the Bureau of Ships.
† Now at Sprague Electric Company, North Adams, Massachusetts.
[1] H. Stohr and W. Klemm, Z. anorg. u. allgem. Chem. 241, 305 (1939).
[2] C. C. Wang and B. H. Alexander, American Institute of Mining and Metallurgical Engineers Symposium on Semiconductors, New York City, February 15–18, 1954 (unpublished). See also Final Technical Report, Bureau of Ships Contract No. NObsr-63180 (unpublished).
[3] Everett R. Johnson and Schuyler M. Christian, Phys. Rev. 95, 560 (1955).
[4] Frank Herman, Phys. Rev. 95, 847 (1954).
[5] Harvey Brooks (private communication).

performed his calculation at a time when the variation of energy gap with composition was not known. He assumed that the energy gap changes linearly with composition of the alloy, and that the change in gap is shared equally by the edges of the valence and conduction band. On this basis Brooks predicted that alloy scattering would be greater than thermal scattering at room temperature, for samples with greater than 8% of minority constituent. If we assume that the temperature dependence of the alloy scattering obtained by Brooks is correct then in the absence of scattering mechanisms other then alloy scattering and thermal scattering, the mobility will be given by

$$\frac{1}{\mu}=\frac{1}{\mu_A}+\frac{1}{\mu_T}=AT^{1.5}+BT^{1.5}. \tag{1}$$

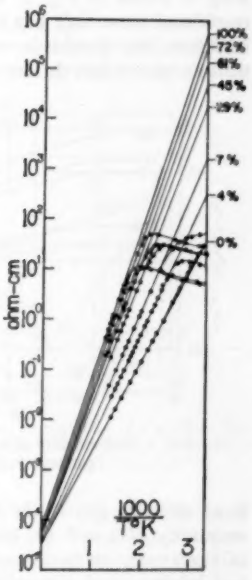


FIG. 1. Resistivity of germanium-silicon alloys *versus* temperature in the intrinsic range. The extrapolation of the intrinsic resistivities to room temperature is shown.

In (1), μ_A is the mobility to be associated with alloy scattering alone and μ_T is the mobility to be associated with thermal scattering alone. For thermal scattering, the exponent 1.5, expected from acoustical mode scattering, is used for illustrative purposes. The argument which follows requires only that the exponent characterizing the thermal contribution to the reciprocal mobility be greater than the exponent characterizing the alloy contribution. With a mobility law of this form it is clear that in the absence of other scattering mechanisms the slope of the log-log plot of mobility against temperature will change from a value characteristic of thermal scattering to one characteristic of alloy scattering as the temperature is reduced. The relative magnitudes of the coefficients A and B, and the magnitude of other scattering mechanisms will determine whether or not alloy scattering is detectable.

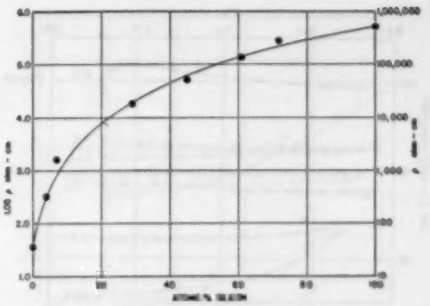


FIG. 2. Intrinsic resistivity of germanium-silicon alloys at room temperature *versus* composition.

The argument presented above suffers from the assumption that thermal scattering in the alloys can be represented as a simple power law in temperature. The calculation of thermal scattering requires that the lattice vibrations be described in normal coordinates. The usual transformation to normal coordinates[6] results in a diagonalization of the configurational energy of the system provided the lattice is periodic. For the alloys, it seems clear that this problem needs to be reexamined before definite statements can be made concerning the thermal scattering.

2. MATERIALS AND EXPERIMENTAL PROCEDURE

The alloys used in this work were prepared by C. C. Wang of this laboratory, using an isothermal solidification technique which effectively avoids the phenomenon of coring and makes it possible to produce homogeneous alloys without annealing. This technique has been previously described.[3] The germanium used was zone purified to room temperature resistivity better than 40 ohm-cm. The silicon was du Pont, classified as hyperpure. No chemical impurity was deliberately added to these alloys. The impurity concentrations indicated by the Hall measurements are of the

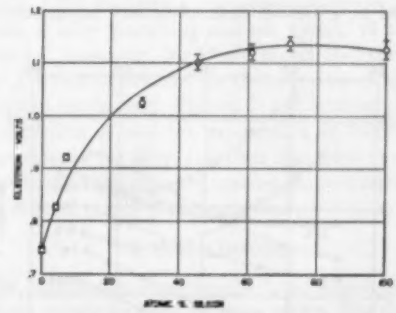


FIG. 3. The energy gap of germanium-silicon alloys *versus* composition.

[6] See, for example, F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), Sec. 22.

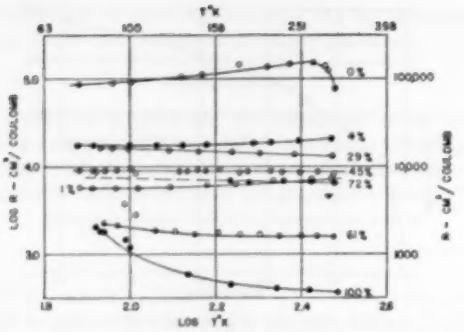


FIG. 4. Hall coefficient *versus* temperature of eight p-type germanium-silicon alloys.

order of 10^{14} per cc. The homogeneity of the alloys was verified by x-ray diffraction. The compositions, indicated in the figures in atomic percent silicon, were determined by chemical analysis. The alloys used in this study are mostly polycrystalline, the exceptions being the 1% and 4% samples which are single crystal.

The electrical measurements performed consisted of resistivity measurements in a temperature range extending from 300°K up to 800°K and Hall and resistivity measurements in a temperature range extending from 77°K up to 300°K. The high-temperature measurements were taken on p-type samples with the exception of the 7% sample which was n-type. These measurements establishes the intrinsic resistivity lines for the alloys. The low-temperature measurements were taken on p-type samples only. These measurements will be shown to yield evidence for the existence of alloy scattering. The magnetic field strength employed for the Hall measurements was 1300 gauss. The measuring technique consisted of the standard potential probe method. High-temperature measurements were performed in an inert atmosphere. Temperature stability for the low-temperature measurements was achieved with the use of a holder of large heat capacity, inserted in a Dewar flask. Auxiliary equipment consisted of appropriate switching circuits and a Leeds and Northrup type K potentiometer.

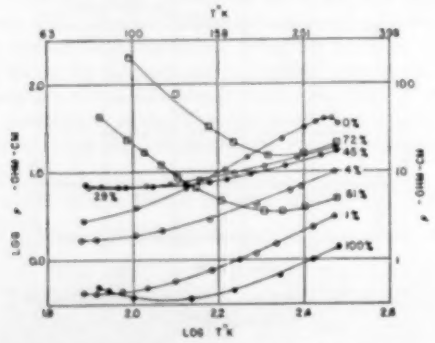


FIG. 5. Resistivity *versus* temperature of eight p-type germanium-silicon alloys.

3. RESULTS AND INTERPRETATION

3.1 Intrinsic Range

The high-temperature resistivity results are shown in Fig. 1, which is a semilog plot of resistivity *versus* reciprocal temperature. The room-temperature intrinsic resistivities, shown in Fig. 2, were obtained from the results of Fig. 1 by extrapolation. It is clear that the intrinsic lines of Fig. 1 can be reconciled by a conventional law of the form:

$$\rho = A \exp(E_G/2kT). \tag{2}$$

The parameter E_G as a function of composition of the alloy is shown in Fig. 3. The parameter E_G may be considered the energy gap for the alloys, although this interpretation should be regarded with caution. If distinct band edges do not exist in the alloys, then E_G

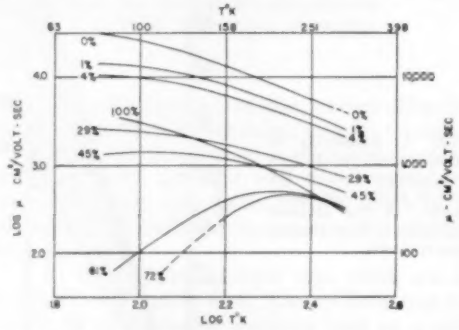


FIG. 6. Hall mobility of holes *versus* temperature in germanium-silicon alloys.

is an effective gap in the sense that it reconciles the resistivity data with the convential law. The variation of energy gap with composition determined optically by Johnson and Christian,[3] exhibits the following characteristics: In the range 0 to 15% silicon, the gap increases linearly from 0.72 to 0.94 ev. A discontinuity in slope then occurs and the gap increases linearly at a lesser rate to a value of 1.20 ev for 100% silicon. Herman,[4] explaining the optical results, has suggested that for an alloy of about 15% silicon the minimum in the conduction band shifts discontinuously in k space. If this is correct, then the resistivity data for alloys near 15% silicon would have to be interpreted in terms of a model which admitted two kinds of carriers in the conduction band and would require two parameters, E_G^1 and E_G^2, for a complete description.

If for the moment Fig. 3 is accepted at face value, that is taken as describing the energy gap of the alloys *versus* composition, it may be noted that the slopes of this curve at the end points are consistent with recent experiments on germanium and silicon at high pressures

reported by Brooks.[7] Brooks has reported that while the coefficient of change of the energy gap of germanium with pressure is positive, the result for silicon indicates a decreasing energy gap with applied pressure. It should be pointed out, however, that the maximum in Fig. 3 is just barely dectable within the experimental accuracy, and may be in error.

The curve of energy gap *versus* composition has been previously published.[8] It is repeated here for the reason that subsequent to the original publication the composition of the alloys were determined more precisely, and this determination required a slight revision of this curve. The energy gap determinations are based on 4 to 8 resistivity runs for each sample.

3.2 Extrinsic Range

Resistivity measurements and Hall measurements at a field strength of 1300 gauss were performed on p-type samples of these alloys in that portion of the extrinsic

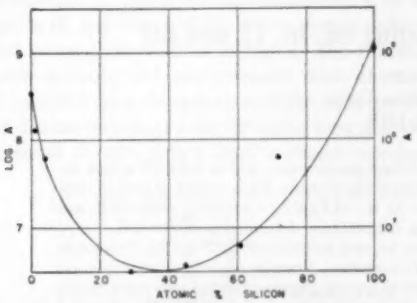


FIG. 7. The coefficient A *versus* composition of the alloy. The Hall mobility of holes in germanium-silicon alloys in the vicinity of 300°K has been reconciled to a law of the form $\mu = AT^{-x}$.

range extending from 77°K to 300°K. The Hall and resistivity data are shown in Fig. 4 and Fig. 5 respectively. These have been combined to give the Hall mobilities for holes, defined as the product of the Hall coefficient and the conductivity, shown in Fig. 6. If one examines the 1%, 4%, 29%, and 45% sample, it may be seen that the scattering mechanisms which reduce the mobilities below the values to be expected from thermal scattering are becoming successively more important and encroach more and more on the thermal scattering range. This is strongly suggestive of alloy scattering. However one would expect the effect to diminish for the 61% and 72% sample. Instead the effect becomes stronger. Moreover the reversal in slope for the 61% and 72% sample is inconsistent with the predicted temperature dependence for alloy scattering. One might attempt to explain these reversals in slope on the basis of impurity scattering. If a comparison is made with work on p-type single crystal silicon, for

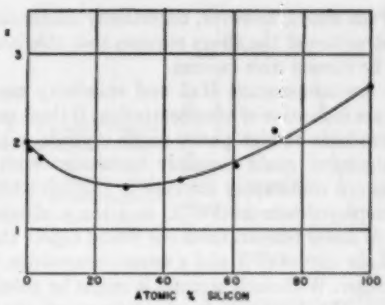


FIG. 8. The exponent x *versus* composition of the alloy. The Hall mobility of holes in germanium-silicon alloys in the vicinity of 300°K has been reconciled to a law of the form $\mu = AT^{-x}$.

instance Morin and Maita's[9] recent work, no such reversals as found in the present work are observed in the same temperature range. If a comparison is made with earlier work on p-type polycrystalline silicon, for instance Pearson and Bardeen's[10] 1949 work, reversals of the sort encountered in the present work are found, but for samples with impurity concentrations several orders of magnitude greater than the impurity concentrations in the samples presently under consideration. While one may ascribe some of the behavior of the 61% and 72% sample to impurity, grain boundary, and alloy scattering, the low-temperature results for these samples must be viewed as anomalous.

It seems clear that if any argument for alloy scattering is to be made for these samples, the argument will have to be based on the results in a temperature range in which impurity and grain boundary effects may be largely neglected. We therefore examine the high-temperature limit of the mobility curves of Fig. 6.

This portion of the curves may be reconciled to a law of the form AT^{-x}. In germanium and silicon it would be the lattice mobility that we would thus represent. In the alloys, if it is true that alloy scattering is important at room temperature, then the exponent x in the law AT^{-x} would be a compromise between a value typical of alloy scattering and one typical of thermal scattering. Moreover, one would expect that the coefficient A would be reduced by the operation of the new scattering mechanism. Figures 7 and 8 show plots of the coefficient A and the exponent x as functions of composition of the alloy. It is clear that these quantities are reduced in the expected manner as one moves away from the end points of these curves.

4. CONCLUSIONS

The high-temperature results have established the intrinsic resistivity lines for these alloys. The slopes of the semilogarithmic plots of resistivity *versus* reciprocal temperature may be interpreted as yielding the energy

[7] Harvey Brooks, Phys. Rev. 98, 244(T) (1955).
[8] Levitas, Wang, and Alexander, Phys. Rev. 95, 846 (1954).
[9] F. J. Morin and J. P. Maita, Phys. Rev. 96, 28 (1954).
[10] G. L. Pearson and J. Bardeen, Phys. Rev. 75, 865 (1949).

gap of the alloys, however, uncertainty concerning the band structure of the alloys requires that this interpretation be viewed with caution.

The low-temperature Hall and resistivity measurements are indicative of alloy scattering. If these samples were available as high-purity single crystals, a conclusive argument could certainly be made. With such samples one could apply the type of analysis which has been employed here at 300°K, to a range of temperatures. At lower temperatures one would expect that the dips in the curves of A and x *versus* composition would be stronger. With such samples it might be possible to establish the temperature dependence of the alloy scattering and to determine the relative magnitudes of alloy and thermal scattering. With the samples presently available, it may be possible to make measurements at somewhat higher temperatures, although not entering the intrinsic range, and to observe a gradual diminuition in the effects of alloy scattering. The curves of A and x would not dip as strongly.

5. ACKNOWLEDGMENTS

The author wishes to express his gratitude to Dr. James F. Battey, Dr. Esther Conwell, and Mr. Robert Baum, all of Sylvania Electric Products, Inc., for their help and encouragement. Special thanks are due to Professor Harvey Brooks of Harvard University for several illuminating discussions on the nature of alloy scattering. The successful conclusion of this work is in large measure due to the help received of Mr. George Servente in performing the electrical measurements.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Effects of Stress on Superconducting Sn, In, Tl, and Al†

NILS L. MUENCH*
The Rice Institute, Houston, Texas
(Received June 15, 1955)

Hydrostatic pressures of 0 to 100 atmos obtained with helium gas pressure, and of 1.9×10^3 atmos obtained by an ice expansion bomb technique are used to measure the pressure displacement of critical temperature T_c. The coefficient $(\partial T_c/\partial p)$ for Sn, In, Tl, and Al is -4.7 ± 0.2, -4.0 ± 0.2, $+0.6\pm0.3$, and -2.0 ± 0.2 in units of 10^{-5} °K/atmos, respectively. The temperature dependence of $R=(\partial H_c/\partial p)_T/(\partial H_c/\partial p)_{T_c}$ above 1°K is described by $R=0.61+0.029T^2$ for Sn, and $R=0.77+0.020T^2$ for In. This result is discussed in connection with the similarity principle, $H_0/T_c=$ constant. From an analysis combining stress and isotope effects, it is seen that T_c, or H_c, is more sensitive to volume changes, holding zero-point lattice vibration amplitude fixed, than to zero-point amplitude changes, holding V fixed: $(\partial H_c/\partial \ln V)_{q^2}=5.5\times(\partial H_c/\partial \ln q^2)_V$ for Sn. Zero-pressure critical field data are also presented for the subject metals.

I. INTRODUCTION

THE dependence of the superconducting critical temperature T_c on mechanical stress[1,2] has received renewed interest since the discovery of the isotope effect and the introduction of the theories by Fröhlich[3] and Bardeen.[4] The critical temperature T_c decreases under pressure for most superconductors studied so far, but increases under pressure for Tl, $NiBi_3$, and $RhBi_4$.[5] As Marcus[6] first noted, the stress effect is not adequately explained as the result of a change in amplitude of the zero-point lattice vibrations. The isotope effect ($M^xT_c=$ constant, where M is the isotopic mass) is usually understood as the result of a change in lattice vibration amplitude. The exponent x is approximately $\frac{1}{2}$, but varies from metal to metal; this has been explained by deLaunay[7] from a consideration of the vibration spectrum of a real lattice. Thus understanding of the isotope effect seems as satisfactory as can be expected in the absence of a good quantum theory of superconductivity. The understanding of stress effects is less satisfactory, and there is clearly a need for more quantitative data. We have studied the displacement of T_c in Sn, In, Tl, and Al using pressures of 0 to 100 atmos obtained with helium gas pressure, and a pressure of 1.9×10^3 atmos obtained by an ice expansion bomb technique.[8] For Sn and In, the value of $(\partial H_c/\partial p)_T$ has been measured as a function of temperature, and it is found to remain finite as T tends to 0°K. The temperature dependence of $(\partial H_c/\partial p)_T$ for tin is shown

† This research was supported in part by the National Science Foundation.
* Magnolia Petroleum Fellow in Physics; now with Humble Oil Research Center, Houston, Texas.
[1] G. J. Sizoo and H. Kamerlingh Onnes, Leiden Comm. No. 180b (1925).
[2] C. F. Squire, *Progress in Low Temperature Physics*, edited by C. J. Gorter (Interscience Publishers, Inc., New York, 1955), Vol. I, p. 151.
[3] H. Fröhlich, Phys. Rev. **79**, 845 (1950).
[4] J. Bardeen, Phys. Rev. **80**, 567 (1950).
[5] Kan, Lazarev, and Sudovstov, J. Exptl. Theoret. Phys. (U.S.S.R.) **18**, 825 (1948); Doklady Akad. Nauk. S.S.S.R. **69**, 173 (1949).
[6] P. M. Marcus, Phys. Rev. **91**, 216 (1953).
[7] J. deLaunay, Phys. Rev. **93**, 661 (1954).
[8] B. G. Lazarev and L. C. Kan, J. Exptl. Theoret. Phys. (U.S.S.R.) **14**, 439 (1944).

to be in agreement with the similarity principle H_0/T_c = constant, but for indium a marked departure from this principle is observed. Finally, an analysis combining stress and isotope data is presented that indicates that T_c, or H_c, is several times more sensitive to fractional changes in volume, holding lattice vibration amplitude fixed, than to fractional changes in vibration amplitude, holding volume fixed.

II. APPARATUS AND TECHNIQUE

The apparatus shown in Fig. 1 is the usual coaxial liquid helium-liquid nitrogen Dewar arrangement. A pressure cell containing a sample 12 cm long and 1 cm in diameter is immersed in the liquid helium. The superconducting transition is observed by means of a drop coil placed in the nitrogen bath and connected to a galvanometer. To obtain the data presented, the pressure and temperature of the specimen were held constant while a vertical magnetic field was increased in small increments. The drop coil was lowered after each change in H, and the resulting galvanometer deflection plotted *vs* magnetic field as in Fig. 2. The transitions for zero pressure and low pressures were about 1% of H_c in width as a consequence of the demagnetizing factor. The transitions at 1.9×10^3 atmos were also about 1 percent of H_c plus only a small constant broadening

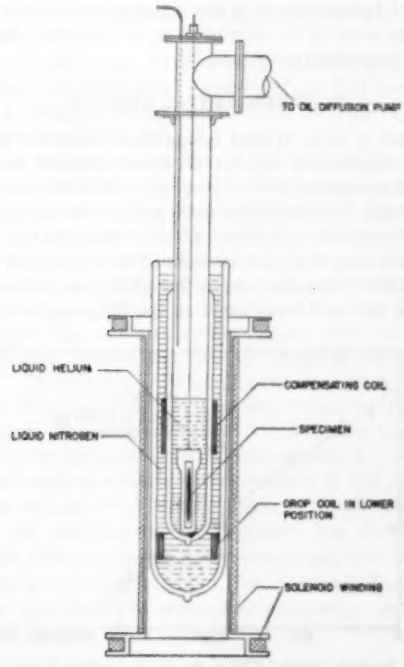


FIG. 1. Experimental apparatus. A radiation shield not shown was installed in the helium Dewar for temperatures in the region of 1°K.

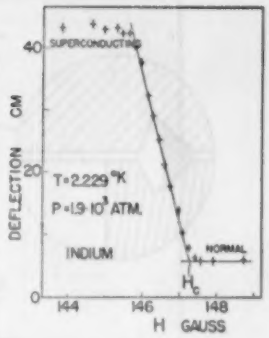


FIG. 2. Typical transition from the superconducting to the normal state at constant temperature and pressure. H_c is determined by extrapolation as shown.

which indicated that the ice pressure was uniform to about 4%.

The solenoid used to produce the vertical field H was calibrated by using a ballistic galvanometer and coil in conjunction with a long thin solenoid. The long solenoid was wound in a single layer in machined grooves so that the magnetic field could be calculated accurately from geometry. Solenoid current was supplied by a generator and electronic regulation system giving a 30-minute stability of 2 parts in 10^4. Current was measured with a potentiometer and standard resistance. The earth's field was compensated by Helmholz coils.

The temperatures above 2°K were controlled through automatic regulation of the vapor pressure. At 3.6°K, the regulation was ±0.00015°K over 30 minute periods. To obtain temperatures between 2°K and 1.3°K the pumping orifice was set and the system was allowed to come to equilibrium with the pumps. To obtain temperatures between 1.3°K and 0.9°K the power input to the MB-200 oil diffusion pump was fixed and the system was allowed to come to equilibrium. In this last temperature range, a spiral radiation shield installed in the helium Dewar reduced the helium boil-off rate by a factor of 3. Vapor pressures were measured with a mercury manometer or McLeod gauge, depending upon pressure. The chief difficulty in temperature control occurs in the He I region where large thermal gradients usually exist in the He bath. Since these gradients can easily introduce an uncertainty of 0.01°K, whereas an accuracy of 0.0002°K is required, considerable attention was devoted to this matter. Originally carbon resistance thermometers with a sensitivity of 0.0001°K were used, but for reasons not understood, the stability of the resistors or measuring circuit over periods of 15 hours was only 0.001°K. Therefore, to obtain most of the data, 10^{-2} watt was supplied to a 500-ohm wire wound resistor placed in the bottom of the bath. This seems to be sufficient power input to insure that the liquid helium is boiling in the vicinity of the specimen. The specimen temperature then was obtained by adding the hydrostatic head pressure of the helium to the measured vapor pressure before referring to the 1949 vapor

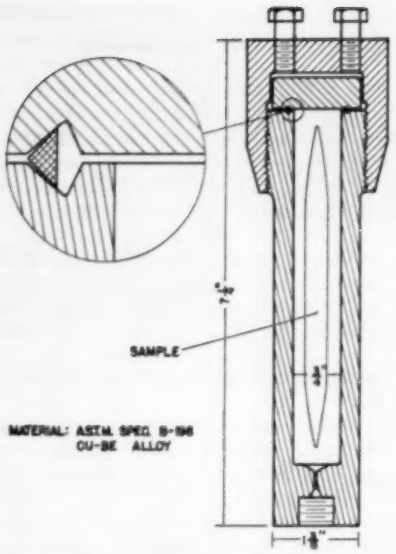


FIG. 3. Nonmagnetic pressure cell for 2×10^9 atmos. This cell is used with the ice technique (1.9×10^9 atmos) and also with helium-transmitted pressures (100 atmos). The seal is made with a triangular cross section metal ring.

pressure tables. To determine the helium level, a number of 160-ohm carbon resistors were strung vertically in the Dewar; a ohmmeter provided sufficient accuracy to determine whether a resistor was immersed in the liquid He.

Two methods were used for applying hydrostatic pressure to the specimen. Pressures of less than 100 atmos were obtained by applying helium gas pressure to a cell containing the specimen. This pressure is limited by the solidification pressure[9] of helium. A pressure of 1.9×10^9 atmos was obtained by freezing water in a constant volume cell (Fig. 3) containing the specimen. From the density of ice I and water, and the thermal expansion and compressibility of the ice, specimen, and cell, one may estimate that a pressure of 3000 atmos would thus be generated. However, a polymorphic phase transition occurs in ice at 2200 atmos.[10] Since this transition is found to proceed rapidly at, say, 240°K, the maximum pressure obtained is fixed by the ice I–ice III equilibrium. If the cell is cooled from 240°K to 78°K in a few hours, the ice I–III and the ice III–II transitions are frozen-in. Then the pressure at 78°K or 4°K is just that of the I–III equilibrium reduced by differential thermal contraction of the ice, specimen, and cell. Lazarev and Kan[8] first used this technique and have studied the matter in detail. We have not repeated their studies of the pressure produced by the ice. Instead we have concluded from the width of superconducting transitions in small fields that a reproducible pressure homogeneous to 4% is obtained. We have used the superconducting transition in tin and indium to determine the pressure produced by the ice bomb. That is, we determine $(\partial T_c/\partial p)$ using helium gas pressure, and then we observe that ΔT_c produced by the ice bomb. Using the fact that $(\partial T_c/\partial p)$ is pressure independent for low pressures,[11] we find that the pressure necessary to produce the observed ΔT_c is 1.9×10^9 atmos for both Sn and In, in fair agreement with Lazarev and Kan and with the pressure expected from the reduced ice I–III equilibrium. For our studies on Al and Tl we have assumed that the ice pressure is again 1.9×10^9 atmos. This assumption is reasonable since the ice bomb pressure is determined primarily by the I–III equilibrium, and since this pressure is found to be the same for Sn and In.

The sources and purities of the specimen metals are as follows: Sn, Vulcan Detinning Company, 99.998%; In, Indium Corporation of America, 99.97%; Tl, Johnson-Mathey Company, Ltd., 99.98%; Al, supplied courtesy of the Aluminum Corporation of America, 99.99%. The Al was received in the form of a cold worked rod and was machined to ellipsoidal shape and annealed at 650°F for 24 hours. The Sn and Tl were cast in a vacuum in a Pyrex tube and cooled slowly from the bottom. The In was drawn from a molten bath of In maintained at the melting temperature. All samples were 12 cm long by 1 cm in diameter, shaped to be approximately ellipsoidal.

III. EXPERIMENTAL RESULTS

Figure 4 gives typical low-pressure data for Sn at fixed temperature. Similar data are obtained for In. H_c is measured as a function of pressure with temperature fixed. To obtain the data points, the pressure is cycled a number of times at each temperature, and several values of H_c are determined each cycle. No cold work effect[12] was observed. Although the vapor pressure is fixed, the bath level, and thus specimen temperature,

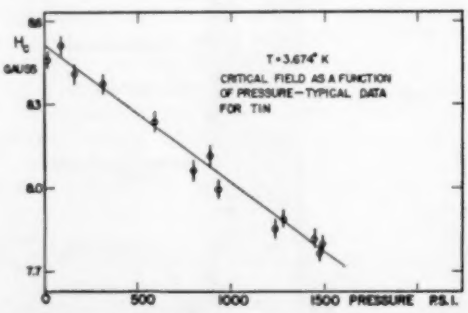


FIG. 4. Typical critical field data for tin at constant temperature.

[9] W. H. Keesom, *Helium* (Elsevier Publishing Company, Amsterdam, 1942).
[10] P. W. Bridgman, Proc. Am. Acad. Arts Sci. 47, 441 (1912).
[11] P. F. Chester and G. O. Jones, Phil. Mag. 44, 1281 (1953).
[12] M. Garbor and D. E. Mapother, Phys. Rev. 94, 1065 (1954).

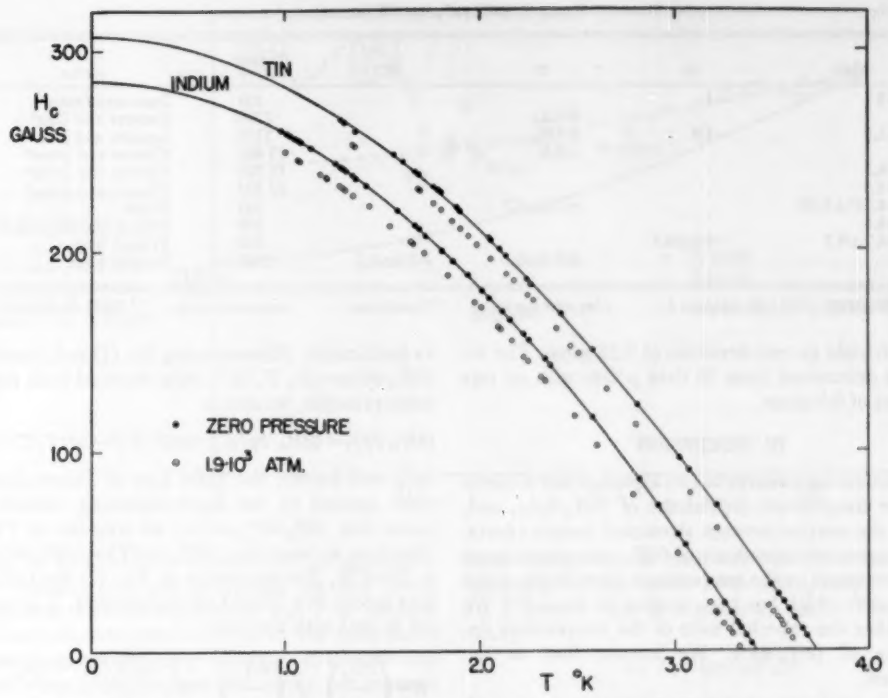


FIG. 5. Critical field data for Sn and In at constant pressure. Zero pressure and 1.9×10^3 atmos are shown.

change with time. Thus a small correction must be applied to the observed values of H_c so that they refer to a common temperature. This is done through knowledge of $(\partial H_c/\partial T)_p$.

The slope of the line through the data of Fig. 4 gives the value of $(\partial H_c/\partial p)_T$ at that temperature. A number of such determinations are made for Sn and for In at temperatures near T_c. The weighted average of these determinations for each metal is given in Table I. $(\partial T_c/\partial p)$ was obtained from $(\partial H_c/\partial p)_T$ through a knowledge of $(\partial H_c/\partial T)_p$. The data for Al and Tl were obtained by determining the critical field curves for both zero pressure and 1.9×10^3 atmos in the vicinity of T_c. The values of T_c were then obtained by extrapolation. The difference in the zero pressure and high pressure T_c was divided by 1.9×10^3 atmos to obtain the values given in Table I. Also included in Table I is a summary of existing determinations of $(\partial T_c/\partial p)$ for these metals.[1,5,11–13] Measurements of simple tension[1,14] are not included because Grenier[15] has shown that stress effects in Sn are highly anisotropic, and thus the pressure coefficient cannot be obtained from tension measurements without a knowledge of crystal orientation.

[13] M. D. Fiske, Phys. Rev. 94, 495 (1954).
[14] Grenier, Spöndlin, and Squire, Physica 19, 833 (1953).
[15] C. Grenier, Compt. rend. 238, 2300 (1954).

The temperature dependence of $(\partial H_c/\partial p)_T$ was determined for Sn and In from the data shown in Fig. 5, which gives the zero pressure and high pressure critical field curves. The temperature dependence is given by specifying the ratio of $\Delta H_c(T)$ to $\Delta H_c(T_c)$, where $\Delta H_c(T)$ is the change in critical field at constant temperature. These data are given in Figs. 6 and 7. The points given were obtained by subtracting each high pressure point from the zero pressure curve adjusted for a best local fit with the zero pressure points. Since Chester and Jones[11] have shown that $(\partial H_c/\partial p)_T$ is independent of pressure to 17 500 atmos, we are justified in replacing $\Delta H_c/\Delta p$ at constant temperature with the more concise notation $(\partial H_c/\partial p)_T$. The least-squares fit for the data in Figs. 6 and 7 gives

$$\left(\frac{\partial H_c}{\partial p}\right)_T \Big/ \left(\frac{\partial H_c}{\partial p}\right)_{T_c} = 0.61+0.029T^2, \quad \text{(Sn)}$$

$$\left(\frac{\partial H_c}{\partial p}\right)_T \Big/ \left(\frac{\partial H_c}{\partial p}\right)_{T_c} = 0.77+0.020T^2. \quad \text{(In)}$$

In the effort to determine small pressure displacements in critical field curves, the zero pressure critical field data were determined with some accuracy. These data are given in Table II. The indium critical field curve is determined from 45 data points from 1°K to

TABLE I. $10^5\times(\partial T_c/\partial p)$(°K/atmos).

Sn	In	Tl	Al	Pressure (atmos)	Author
−3	−4			100	Sizoo and Onnes[a]
		+0.6		1370	Lazarev and Kan[b]
−5.8	−4.8	+1.4		1750	Lazarev and Kan[b]
		−0.4		13 400	Chester and Jones[c]
−4.3				11 500	Chester and Jones[c]
−4.5				17 500	Chester and Jones[c]
−4.58[d]±0.09		+1.3±0.2		100	Fiske[e]
−4.4				100	Garbor and Mapother[f]
−4.7±0.2	−4.0±0.3			100	Present work
		+0.6±0.3	−2.0±0.2	1900	Present work

[a] See reference 1. [b] See reference 5. [c] See reference 11. [d] Unpublished. [e] See reference 13. [f] See reference 12.

T_c which yield an rms deviation of 0.25 gauss. The tin curve is determined from 38 data points with an rms deviation of 0.9 gauss.

IV. DISCUSSION

We confine our remarks to two aspects of our subject: first, the temperature dependence of $(\partial H_c/\partial p)_T$, and, second, the relation between stress and isotope effects. The temperature dependence of $(\partial H_c/\partial p)_T$ is analogous in some respects to the temperature dependence of the isotope shift which has been studied in detail.[16,17] We first outline the expected form of the temperature dependence of $(\partial H_c/\partial p)_T$. We assume that in the expression,

$$H_c(T)=H_0[1+\alpha(T/T_c)^2+\beta(T/T_c)^4], \qquad (1)$$

the ratio H_0/T_c and the coefficients α and β are independent of pressure. This assumption constitutes what is called the "similarity principle"; we will later discuss its justification. Differentiating Eq. (1) and substituting $(\partial H_0/\partial p)_T=(H_0/T_c)(\partial T_c/\partial p)_T$ obtained from the similarity principle, we obtain

$$(\partial H_c/\partial p)_T=(\partial H_0/\partial p)_T[1-\alpha(T/T_c)^2-2\beta(T/T_c)^4]. \qquad (2)$$

As is well known, the Third Law of Thermodynamics when applied to the superconducting transition requires that $(\partial H_c/\partial T)_p=0$ for all pressures at $T=0$°K. Therefore we have that $(\partial^2 H_c/\partial p\partial T)=(\partial^2 H_c/\partial T\partial p)=0$ at $T=0$°K. The expression of Eq. (2) for $(\partial H_c/\partial p)_T$ does satisfy this Third Law requirement. It is convenient to deal with the ratio

$$R=\left(\frac{\partial H_c}{\partial p}\right)_T\Big/\left(\frac{\partial H_c}{\partial p}\right)_{T_c}=\frac{1-\alpha(T/T_c)^2-2\beta(T/T_c)^4}{1-\alpha-2\beta}.$$

The solid curves in Figs. 6 and 7 represent this ratio with the values of α and β determined by least squares fit with zero pressure data. We observe from Fig. 6

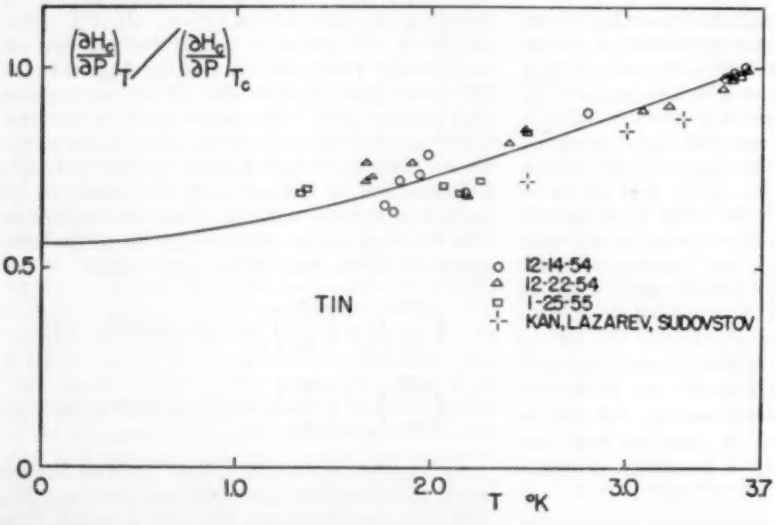


FIG. 6. Temperature dependence of $(\partial H_c/\partial p)_T$ for Sn. The solid line is derived from the similarity principle. The experimental apparatus was disassembled, reassembled, and zero pressure data were taken before and after the dates indicated for pressure data.

[16] Lock, Pippard, and Shoenberg, Proc. Cambridge Phil. Soc. 47, 811 (1951).
[17] E. Maxwell, Phys. Rev. 86, 235 (1952).

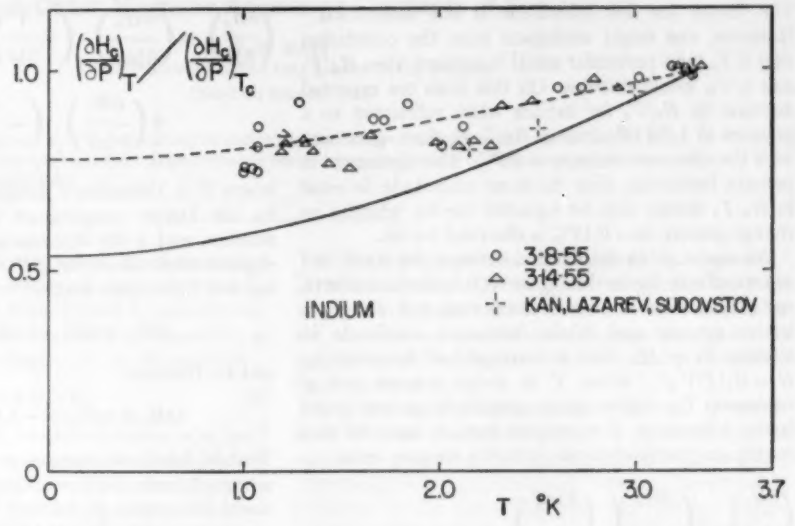


FIG. 7. Temperature dependence of $(\partial H_c/\partial p)_T$ for In. The solid line is derived from the similarity principle. The dashed line is derived by assuming the similarity principle in error by 0.7 percent at 1.9×10^9 atmos.

that for tin the solid curve obtained on the assumption of the similarity principle agrees with the data within experimental error. However, from Fig. 7 we see that for indium all data points lie above the solid curve obtained from the similarity principle. It is clear that the similarity principle does not apply for pressure effects in indium. The dashed curve through the data points in Fig. 7 was obtained on the assumption that α and β are constant, but that H_0/T_c decreases by 0.7% under the ice pressure of 1.9×10^9 atmos. The volume change produced by this pressure is 0.35% for tin and 0.7% for indium.

To review the theoretical basis for the similarity principle (H_0/T_c=constant), we write the Gibbs potential based on the Gorter-Casimir two fluid model.[18] We assume an energy gap Δ per unit volume between normal and superconducting electron energy levels and a fraction $f=n_s(T)/n_s(0°\text{K})$, where $n_s(T)$ is the number of superconducting electrons per unit volume at $T°$K:

$$G_s=U_0-(V/V_m)(1-f)^x\tfrac{1}{2}\gamma T^2+\Delta Vf+G(T)_{\text{lattice}}.$$

Here γT is the electronic specific heat per mole, V_m is molar volume, and x is chosen to fit experiment. For example, $x=\frac{1}{2}$ for a parabolic curve of $H_c(T)$ *vs* T. For stable equilibrium we have $\partial G_s/\partial f=0$. Performing the differentiation, we obtain $f=1-(\Delta V_m/x\gamma T^2)^{1/(x-1)}$. At T_c, f is zero. Also Δ is given by $\frac{1}{2}\mu_0 H_0^2$. We obtain finally:

$$H_0/T_c=(x\gamma/\mu_0 V_m)^{\frac{1}{2}}. \tag{3}$$

This is the similarity principle providing $x\gamma/V_m$ is constant. The exponent x can be expressed as a function of α and β only,[17] and thus its pressure dependence is expected to be small. The electronic specific heat γ can be expressed as the free-electron specific heat by replacing the electron mass with an effective mass m^*. Thus, we have

$$\gamma/V_m=3.26\times10^{-5}(m^*/m)(n^{\frac{1}{3}}V_m^{\frac{1}{3}}/V_m)\ \text{cal/deg-mole}^2,$$

where n is the number of valence electrons per atom. To explain the observed dependence of H_0/T_c on volume in the case of indium one must assume that m^* is approximately proportional to V_m^2.

From another standpoint, we have empirical evidence for the dependence of H_0/T_c or γ/V_m on T_c. The value of H_0/T_c or γ/V_m for each superconductor can be plotted as a function of T_c. One finds that the soft superconductors fall approximately along a straight line (with a branch line below 2°K) with positive slope.

TABLE II. Zero-pressure data.

Metal	T_c(°K)	$(\partial H_c/\partial T)_{T=T_c}$ (gauss/°K)	$H_c=H_c(T)$ (gauss)
Sn	3.728±0.0015	−147.4±1	$H_c=307.3-26.80T^2+1.225T^3$ $=H_0[1-1.2117(T/T_c)^2+0.2117(T/T_c)^3]$
In	3.412±0.003	−155.8±1	$H_c=284.5-27.68T^2+0.9489T^3$ $=H_0[1-1.1325(T/T_c)^2+0.1325(T/T_c)^3]$
Tl	2.385±0.004	−135.5±2	...
Al	1.174±0.003	−163±3	...

[18] P. L. Bender and C. J. Gorter, Physica 18, 597 (1952).

The reason for this correlation is not understood.[19] However, one might anticipate from the correlation that if T_c for a particular metal is reduced, then H_0/T_c and γ/V_m would decrease. On this basis the expected decrease in H_0/T_c for indium when submitted to a pressure of 1.9×10^9 atmos is 0.6%, in close agreement with the observed decrease of 0.7%. This agreement is perhaps fortuitous, since the same magnitude decrease in H_0/T_c should then be expected for Sn, whereas no change greater than 0.15% is observed for tin.

We now turn to the relation between the stress and isotope effects. By combining the data from these effects, one may estimate the relative importance of changes in lattice spacing and lattice vibration amplitude in altering T_c or H_0. This is accomplished by assuming $H_c=H_c(T,V,q^2)$, where V is molar volume and q^2 represents the mean square amplitude of zero point lattice vibrations. If we regard isotopic mass M as a continuous thermodynamic variable, we may write

$$\left(\frac{\partial H_c}{\partial p}\right)_T=\left(\frac{\partial H_c}{\partial \ln q^2}\right)_V\left(\frac{\partial \ln q^2}{\partial p}\right)_T+\left(\frac{\partial H_c}{\partial \ln V}\right)_{q^2}\left(\frac{\partial \ln V}{\partial p}\right)_T \quad (4)$$

$$\left(\frac{\partial H_c}{\partial M}\right)_T=\left(\frac{\partial H_c}{\partial \ln q^2}\right)_V\left(\frac{\partial \ln q^2}{\partial M}\right)_T+\left(\frac{\partial H_c}{\partial \ln V}\right)_{q^2}\left(\frac{\partial \ln V}{\partial M}\right)_T. \quad (5)$$

The left members are measured experimentally. The coefficients of $(\partial H_c/\partial \ln q^2)_V$ and $(\partial H_c/\partial \ln V)_{q^2}$ can be estimated, with the result that Eqs. (4) and (5) become

$$\left(\frac{\partial H_c}{\partial p}\right)_T=\left(\frac{\partial H_c}{\partial \ln q^2}\right)_V(-\Gamma B)+\left(\frac{\partial H_c}{\partial \ln V}\right)_{q^2}(-B) \quad (4A)$$

[19] J. G. Daunt, *Progress in Low Temperature Physics*, edited by C. J. Gorter (Interscience Publishers, Inc., New York, 1955), Vol. I, p. 202.

$$\left(\frac{\partial H_c}{\partial M}\right)_T=\left(\frac{\partial H_c}{\partial \ln q^2}\right)_V\left(-\frac{1}{2M}\right)+\left(\frac{\partial H_c}{\partial \ln V}\right)_{q^2}\left(-\frac{1}{2M}\right)\left(\frac{9}{8}\frac{Nk\theta_D\Gamma}{V}\right), \quad (5A)$$

where Γ is Grüneisen's constant,[20] B compressibility, θ_D the Debye temperature at 0°K, N Avogadro's number, and k the Boltzmann constant. If we insert experimental values for $(\partial H_c/\partial p)_T$, $(\partial H_c/\partial M)_T$, B, Γ, θ_D, and V, we have for tin:

$$(\partial H_c/\partial \ln V)_{q^2}=+5.5(\partial H_c/\partial \ln q^2)_V,$$

and for thallium:

$$(\partial H_c/\partial \ln V)_{q^2}=-3.3(\partial H_c/\partial \ln q^2)_V.$$

That is, fractional changes in V, holding q^2 constant, are much more effective in displacing H_c than are fractional changes in q^2, holding V constant. Alternately, one may consider $H_c=H_c(T,V,\theta_D)$. In this case similar expressions are obtained, with the result that $(\partial H_c/\partial \ln V)_{\theta_D}=10(\partial H_c/\partial \ln\theta_D)_V$ for tin.

Since the energy gap Δ between normal and superconducting energy levels on the two-fluid model is given by $\frac{1}{2}\mu_0 H_c^2$, this analysis seems to indicate that the energy gap Δ, and therefore the interaction mechanism responsible for superconductivity, is more sensitive to changes in lattice dimension than to zero-point lattice vibration amplitude.

V. ACKNOWLEDGMENTS

The author is indebted to Professors C. F. Squire and H. E. Rorschach, Jr., for stimulating advice and encouragement. The author wishes to express his appreciation to the members of the Low Temperature Laboratory of the Rice Institute for assistance in obtaining the data presented.

[20] J. C. Slater, *Introduction to Chemical Physics* (McGraw-Hill Book Company, Inc., New York, 1939), p. 219.

First-Order Magnetic Anisotropy Constants of Ferrites

J. B. Birks
British Dielectric Research Ltd., London W. 12, England
(Received May 26, 1955)

Certain sources of error in attempting to derive the first order anisotropy constants of ferrites from magnetostatic data are pointed out. High-frequency observations, particularly of the natural spin resonance frequency, are preferable.

WEISZ[1] has recently reported quasi-theoretical values of the first-order anisotropy constant K_1 of several ferrites. These were derived by substitution of experimental values of the initial permeability μ_0 and the saturation magnetization M_s in the equation

$$\mu_0 - 1 = kM_s^2/K_1. \quad (1)$$

The constant k, for which theoretical values of π, $4\pi/3$ and $8\pi/3$ have been variously proposed, is obtained by substitution in (1) of experimental data for magnetite ($\mu_0=45$, $M_s=480$, $K_1=-1.1\times10^5$ ergs/cm^3). This gives $k=21$, and the same value is assumed for the other ferrites. The purpose of the present note is to point out several potential sources of error in this procedure, and to mention some earlier published work which appears to have been overlooked.

1. The quantity μ_0 in (1) refers to the component of the initial permeability due to spin rotation only, and not to the total static initial permeability μ_s, the experimental values of which are used by Weisz.

2. Although Birks,[2] and later Went and Wijn,[3] have shown spin rotation to be the principal magnetization process in several low-conductivity ferrites, the difference between μ_s and μ_0 is not always negligible; for example, for Ni–Zn ferrite, $\mu_s=40$, $\mu_0=34$.[2] The error introduced by equating μ_0 and μ_s can be eliminated by permeability measurements at frequencies of ~200 Mc/sec, between the dispersion regions due to translational (domain wall displacement) and rotational processes.

3. The error is particularly significant in magnetite,[2] in which $\mu_s\sim2\mu_0$. The magnetic dispersion[4] at 5–20 Mc/sec in this material is attributed to relaxation of the translational magnetization component, and it is only at higher frequencies that the initial permeability becomes equal to μ_0.

4. The translational component $(\mu_s-\mu_0)$ is associated with the conductivity, which is relatively high in magnetite compared with the other ferrites. μ_s is therefore susceptible to impurities and strains. This probably accounts for the major difference between the values for magnetite powder of $\mu_s=15$, $\mu_0=8$, observed by Birks,[2] and the value for sintered magnetite of $\mu_s=45$, used by Weisz.[1] The former experimental data substituted in (1) give a value of $k=3.8$, which lies within the range of theoretical values of this constant.

5. An alternative method of determining $|K_1|$ for powdered ferrites by observations of the natural spin resonance frequency in the internal anisotropy field has been described previously.[2] This method gives a value of $|K_1|=1.3\times10^5$ ergs/cm^3 for magnetite, compared with that of $K_1=-1.1\times10^5$ ergs/cm^3 derived[5] from the single-crystal data of Quittner,[5] and subsequently confirmed by Bickford[6] and Danan.[7]

There are insufficient data available to estimate the resultant errors in the absolute values of $|K_1|$ derived by Weisz,[1] but it is unlikely that they are of sufficient magnitude to modify the broad qualitative correspondence found between $|K_1|$ and J, the inner quantum of the M^{++} ions in the ferrite structure, $MO\cdot Fe_2O_3$.

[1] R. S. Weisz, Phys. Rev. **96**, 800 (1954).
[2] J. B. Birks, Proc. Phys. Soc. (London) **B63**, 65 (1950).
[3] J. J. Went and H. P. J. Wijn, Phys. Rev. **82**, 269 (1951).
[4] J. Zimowski, Acta Phys. Polon. **6**, 6 (1937).
[5] V. Quittner, Ann. Physik **30**, 289 (1909).
[6] L. R. Bickford, Phys. Rev. **78**, 449 (1950).
[7] H. Danan, Compt. rend. **238**, 1304 (1954).

Concept of Temperature and the Overhauser Nuclear Polarization Effect

CHARLES P. SLICHTER
Department of Physics, University of Illinois, Urbana, Illinois
(Received May 26, 1955)

The polarization of nuclei by the Overhauser method is examined for the case of metals. It is shown that the concept of temperature is valid for the conduction electrons if they are viewed form a coordinate system rotating at the electron Larmor frequency ω. Relative to this system, the nuclei precess at nearly $-\omega$, a result equivalent to an enormous nuclear magnetic coupling. The Overhauser result then follows immediately.

OVERHAUSER[1] has proposed a method for polarizing nuclear spins. It has been verified experimentally both in metals[2] and in nonmetals.[3,4] Prior to the experimental verification, which actually preceded the publication of Overhauser's detailed account, there was some skepticism expressed, one argument running as follows.

The nuclei become polarized by interaction with an electron system which is held saturated by a strong alternating magnetic field. By saturation we mean equal numbers of electron spins in the two spin orientations. Since a saturated magnetic resonance corresponds to an infinite spin temperature, the electron spin system has an infinite temperature. Therefore, the nuclei must arrive at infinite spin temperature since their contact with their surroundings is via the electrons. Hence, the nuclear polarization is *zero*.

The above argument, though plausible, is incorrect. As we shall discuss, the electrons cannot be said to have infinite temperature, because their distribution function is not *describable* by a temperature. If we could describe the electrons by a temperature, we could readily derive the nuclear polarization as the thermal equilibrium value. The purpose of this paper is to show that the temperature concept applies to the electrons in a properly chosen rotating coordinate system, relative to which the apparent nuclear Zeeman splitting is very much larger than usual. The resulting thermal equilibrium nuclear polarization is just that predicted by Overhauser. Let us turn to the details.

The Hamiltonian describing the conduction electron and nuclear system may be written as

$$\mathcal{H}=\mathcal{H}_T+\mathcal{H}_{SL}+\gamma_e\hbar H_0\sum_i S_{zi}+\gamma_n\hbar H_0\sum_j I_{zj}$$
$$+\frac{8\pi}{3}\gamma_e\gamma_n\hbar^2\sum_{i,j}\delta(\mathbf{r}_i-\mathbf{r}_j)\mathbf{S}_i\cdot\mathbf{I}_j+\mathcal{H}_{\rm sat},$$

where $\mathcal{H}_T$ represents the electron kinetic and potential energies, $\mathcal{H}_{SL}$ is the electron spin-lattice interaction, γ_e and γ_n the electron and nuclear gyromagnetic ratios, H_0 the applied static magnetic field (in the z-direction), $\mathbf{S}$ and $\mathbf{I}$ the electron and nuclear spin angular momenta, and $\mathcal{H}_{\rm sat}$ the interaction which saturates the electron spin system.

The terms $\mathcal{H}_{SL}$ and $\mathcal{H}_{\rm sat}$ determine the relative population of electron spins up or down. We consider the case of complete saturation, in which case there are equal numbers of electrons spins up or down. We will then neglect the presence of the terms $\mathcal{H}_{SL}$ and $\mathcal{H}_{\rm sat}$, and simply consider that the electron spin population is maintained saturated despite the electron-nuclear couplings.

If we think of the electrons as described by two Fermi distributions, one for each spin orientation, we can see readily why the concept of electron temperature fails. The bottoms of the distributions, representing zero electron kinetic energy, are displaced by the electron magnetic energy. Since there are equal numbers of electrons in both distributions the tops are also displaced by the same amount. Clearly, there is *no* temperature which corresponds to a displacement of the tops. If the tops were in coincidence, the temperature concept would apply, and the temperature would be simply the "lattice" temperature T, which describes the shape of the tail.

If we now transform to a reference system rotating at the electron Larmor frequency, we cancel off the electron magnetic energy. That is, speaking classically, in the rotating reference system the electron's precessional motion has been "stopped." Mathematically we produce the transformation by the operator $\exp(-i\gamma_e\hbar H_0\sum_j S_{zj})$, which removes the electron magnetic energy from $\mathcal{H}$. However, this transformation would introduce explicit time dependence into the $\mathbf{I}\cdot\mathbf{S}$ coupling term, since $\sum_j S_{zj}$ does not commute with this operator. Explicit time dependence in a Hamiltonian corresponds to a "driven" system, so that energy conservation no longer applies within the nuclear plus electron system. To avoid considering a nonconservative system, therefore, we use instead the operator $\exp[-i\gamma_e\hbar H_0\sum_j(S_{zj}+I_{zj})]$ which rotates *both* electrons and nuclei at the same rate, $\gamma_e\hbar H_0$, and leaves the coupling term invariant.

[1] A. W. Overhauser, Phys. Rev. **91**, 476 (1953); **92**, 411 (1953). For further discussions see, for example, F. Bloch, Phys. Rev. **93**, 944(A) (1954); C. Kittel, Phys. Rev. **95**, 589 (1954); J. Korringa, Phys. Rev. **94**, 1388 (1954); A. Abragam, Phys. Rev. **98**, 1729 (1955).
[2] T. R. Carver and C. P. Slichter, Phys. Rev. **92**, 212 (1953).
[3] T. R. Carver, thesis, University of Illinois, 1954 (unpublished).
[4] Beljers, Van der Kint, and Van Wieringen, Phys. Rev. **95**, 1683 (1954).

The transformed Hamiltonian is then

$$\mathcal{H}=\mathcal{H}_T+(\gamma_e\hbar H_0-\gamma_e\hbar H_0)\sum_i S_{zi}$$
$$+(\gamma_n\hbar H_0-\gamma_e\hbar H_0)\sum_j I_{zj}+\frac{8\pi}{3}\gamma_e\gamma_n\hbar^2\sum_{i,j}\mathbf{S}_i\cdot\mathbf{I}_j.$$

We note that the electron Zeeman energy is now zero. The electrons are now describable by a temperature since the tops of the Fermi surfaces coincide, but the nuclei have acquired a different splitting, equivalent to their having a gyromagnetic ratio nearly that of the electron. Since the temperature concept applies to the electrons, we can say the nuclei will come to thermal equilibrium at temperature T, among levels of spacing $(\gamma_n-\gamma_e)\hbar H_0$. Thus the Overhauser polarization is produced.

We can summarize, then, by saying that to apply the concept of temperature we must view the nuclei and electrons from a system rotating with the electrons. Relative to this system, the nuclei are precessing at nearly the electron Larmor frequency. Such a rapid precession is equivalent to the production of an enormous nuclear magnetic interaction.

As a subsidiary consideration, we might point out that in many ways the Overhauser effect has been with us a long time, although unnoticed. Let us think of the magnetization of a paramagnetic sample in a static field. The spin polarization is determined by the Boltzmann exponent $\gamma\hbar H_0/kT$. If we now transform to a rotating reference system to cancel off the magnetic energy, we realize there is a seeming contradiction since the full polarization must still be produced although the magnetic interaction is zero. The explanation of our difficulty is clearly that the lattice (which we must rotate to avoid explicit time dependence in the spin-lattice coupling) is no longer described by the temperature T. That is, the spacing of lattice energy levels has collapsed, giving a much greater population difference for the (zero) energy gap than we would get if we used the energy gap in the rotating system and the temperature T.

Our considerations should not be confused with the interesting discussion by Redfield[6] of the use of the temperature concept with rotating reference frames to discuss saturation. He is concerned with the coherent transverse magnetization of a saturated spin system with strong spin-spin interactions. We are concerned with the longitudinal magnetization of the nuclear spin system produced by coupling to a saturated spin system (that of the electron) of greatly different Lamor frequency, and possessing additional (translational) degrees of freedom. We have neglected effects of the sort described by Redfield.

[6] A. G. Redfield, Phys. Rev. **98**, 1787 (1955).

Gyromagnetic Ratio of Nickel at Low Magnetic Intensities

G. G. Scott
Research Laboratories Division, General Motors Corporation, Detroit, Michigan
(Received May 31, 1955)

It has recently been shown that the measured value for the gyromagnetic ratio of iron, as measured by an Einstein-DeHaas experiment, is field-dependent in the low-intensity region, and that the extrapolated zero-intensity value of this ratio checks the value which is theoretically expected from a consideration of recent ferromagnetic resonance experiments.

This experiment has been repeated for a sample of pure nickel and a similar result has been obtained. The extrapolated zero intensity value for g' for nickel is 1.801. At a higher induced magnetic intensity, the effective value of g' changes to 1.830.

INTRODUCTION

THE techniques used in these experiments have been previously described.[1,2] The rod used was of high-purity nickel, very kindly furnished by the International Nickel Company. It is the same rod that was previously used to measure the gyromagnetic ratio of nickel at the General Motors Research Laboratories. In its original state, this rod was hot-rolled, annealed, and then cold drawn 40%.

After making the improvements in the equipment described in reference 2, a series of check readings were taken on the Ni rod. These readings were taken with the rod in its original state on 10 different days during December, 1953. The value obtained from this set of runs was $\rho e/m=1.092$ or $g'=1.831$. These runs were taken using a magnetizing current of 16.00 milliamperes. The resulting magnetic moment of the rod alone was 20 871 amp cm².

After the above runs were completed, the winding was removed and the rod was annealed for 2 hours in dry hydrogen at 2000°F and furnace cooled. This resulted in a fourfold increase in magnetic permeability. The rod was rewound and the present series of experiments on the field dependence of g' were performed with the rod in this annealed condition.

TABLE I. Condensed data for determining gyromagnetic ratios of nickel.[a]

$$\rho e/m=\left(\frac{\pi Id}{4PXkm/e}-2i_s\Sigma A_c\right)\Big/(M_s-i_s\Sigma A_c).$$

i_s(ma)	M_s	P	d	$\rho e/m$
4.0008	21814	28.117	0.036084	1.097
4.0001	21811	28.105	0.036044	1.097
2.0000	10730	28.105	0.017940	1.110
2.0000	10730	28.107	0.017941	1.110
2.0000	10730	28.107	0.017951	1.110
8.0034	43902	28.122	0.072348	1.093
8.0011	43890	28.122	0.072480	1.095
16.0030	86570	28.154	0.14251	1.090
2.0001	10730	28.100	0.017963	1.111
1.0000	5203	28.109	0.008708	1.110
1.0000	5203	28.107	0.008671	1.105
1.0000	5203	28.106	0.008711	1.111
1.0000	5203	28.106	0.008691	1.108
1.0000	5203	28.105	0.008713	1.111
1.0000	5203	28.105	0.008731	1.113
3.0008	16247	28.103	0.027031	1.104
3.0004	16244	28.107	0.026958	1.101
5.0004	27326	28.109	0.045194	1.097
5.0004	27326	28.109	0.045266	1.099
4.0002	21811	28.108	0.036168	1.100
4.0004	21812	28.108	0.036126	1.099
3.0001	16243	28.106	0.027004	1.103
3.0000	16242	28.101	0.027086	1.107

[a] Moment of inertia, $I=215.26$ g cm².
Optical length, $X=1576.9$ cm.
Phase angle constant, $k=0.99941$.
Winding constant, $\Sigma A_c=78490$ cm².
Mass charge ratio of electron, $m/e=5.6844\times10^{-9}$ g coul⁻¹.
i_s = magnetizing current, amperes.
M_s = magnetic moment (rod and winding), amp cm².
P = period, seconds.
d = amplitude change per reversal of i_s, cm.

RESULTS

The condensed data shown in Table I were taken during March, and April, 1955. The average value of $\rho e/m$ for each value of the magnetizing current used is given in Table II along with the corresponding values of g'. Figure 1 is a plot of the data given in Table II. The average induced magnetic intensity of the rod which is plotted along the X axis is obtained by dividing the magnetic moment of the rod alone in amp cm² by the volume of the rod.

It can be seen from an examination of this curve that the change from the low intensity value for g' to the high intensity value apparently takes place more abruptly than is the case for Fe.

It seems possible that this change in g' might occur at a unique value of the induced magnetic intensity and that the gradual change obtained is the result of

TABLE II. Values of $\rho e/m$ and g' for various induced magnetic intensities in nickel.[a]

i_s	$\mathfrak{B}$	$\rho e/m$	g'
1.0	131.8	1.110	1.802
2.0	271.9	1.110	1.801
3.0	411.8	1.104	1.812
4.0	552.9	1.098	1.821
5.0	692.7	1.098	1.821
8.0	1112.8	1.094	1.828
16.0	2194.3	1.090	1.834

[a] i_s = winding current, milliamperes.
$\mathfrak{B}$ = average induced magnetic intensity of rod, amp cm⁻¹.
$\mathfrak{B}=(M_s-i_s\Sigma A_c)/v$. $v=38.88$ cm³.

[1] G. G. Scott, Phys. Rev. 82, 542 (1951).
[2] G. G. Scott, Phys. Rev. 99, 1241 (1955).

nonuniformity of magnetization of the rod. If it is assumed that the rod is elliptically magnetized and that g' changes from 1.801 to 1.830 at an induced intensity of 470 amp cm^{-1}, then a curve of g' *vs* average induced intensity would be obtained, which would be nearly the same as the experimental curve. In this case, however, there would be a discontinuity in the curve at 370 amp cm^{-1} which is the average intensity at which the central region of the rod reaches the assumed critical intensity of 470 amp cm^{-1}. This hypothetical discontinuity is shown as a broken line in Fig. 1. More points and a winding which more uniformly magnetizes the

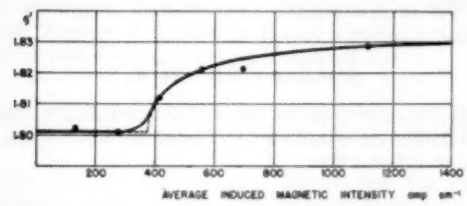


FIG. 1. Plot of g' *vs* average induced magnetic intensity for Ni rod 1.5-cm diam, 22 cm long.

specimen would undoubtedly be needed in order to decide whether or not this change occurs at a unique value of the magnetic intensity.

The extrapolated zero intensity value of 1.801 for g' is seen to check very well the value expected from microwave experiments.[3]

[3] C. Kittel, *Introduction to Solid State Physics* (John Wiley and Sons, Inc., New York, 1953), p. 171.

TABLE III. Chemical analysis of impurities in nickel rod.[a]

Iron	0.032%
Cobalt	Negligible
Manganese	0.030%
Copper	0.003%
Silicon	0.10%
Carbon	0.01%
Sulfur	0.005%

[a] Analysis made by International Nickel Company.

The higher intensity value of 1.830 checks work which has previously been done on Ni at the General Motors Research Laboratories.[4] A chemical analysis of the Ni rod is given in Table III.

CONCLUSION

It has been shown from an Einstein-DeHaas experiment that the effective value of g' for Ni depends on the induced magnetic intensity of the specimen.

When the curve of g' *vs* induced intensity is extrapolated to zero intensity, the value obtained for g' is 1.801 ± 0.002. A similar effect has recently been reported for Fe.

ACKNOWLEDGMENTS

The author wishes to thank the members of the staff of the General Motors Research Laboratories who have cooperated in this experiment. Thanks are also due the International Nickel Company for supplying the nickel rod.

[4] Brown, Meyer, and Scott, Compt. Rend. **238**, 2504 (1954).

Computation of Mean Debye Temperature of Cubic Crystals from Elastic Constants. II

PAUL M. SUTTON
Corning Glass Works, Corning, New York
(Received February 28, 1955)

Quimby and Sutton have presented, in an array of equations, an improved, accurate method for computing the mean Debye characteristic temperature of cubic crystals from the elastic constants. The present paper gives an equivalent, simplified, explicit equation. The pertinent integral is also expanded in a Taylor series to the fifth term by a special device and a method of getting the general term made evident. The first two terms of this series constitute Born's equation which is shown to be usually inaccurate since the series is slowly convergent. This series is useful, however, in estimating the correction to the simple equation first mentioned. This correction is analyzed in more detail and extended to values of the anisotropy factor less than unity. Graphs are given permitting rapid calculation of Debye temperatures with an uncertainty due to the approximation made in the method of about 0.1 percent.

INTRODUCTION

IN a previous paper by Quimby and Sutton,[1] a method was presented for computing the mean Debye characteristic temperature, Θ_D, of cubic crystals from the elastic constants. It involves the evaluation of an integral

$$Y=\int_0^{4\pi}\left[\sum_i \frac{1}{(C+z_i)^3}\right]^{\frac{1}{3}}\frac{d\Omega}{4\pi}, \qquad (1)$$

where $C=c_{44}/(c_{11}-c_{44})$, c_{jk} are the elastic constants, $d\Omega$ is an element of solid angle, and the z_i are the three real roots of a cubic equation. The z_i depend on the direction cosines for a given direction, α, β, γ, and on $K=(c_{12}+c_{44})/(c_{11}-c_{44})$.

An approximate formula for Y, denoted Y_R, is devised by Quimby and Sutton[1] by a modification of the method of Hopf and Lechner.[2] The function $f(z)=(C+z)^{-\frac{1}{3}}$ is represented by a fifth-degree polynomial $[f(z)]_R$ given by the expression

$$[f(z)]_R=a_0+a_1z+a_2z^2+a_3z^3+a_4z^4+a_5z^5, \qquad (2)$$

and the a_i are so adjusted that the value of $[f(z)]_R$ coincides with that of $f(z)$ at six appropriately chosen values of z. Formulas are given for the a_i and a formula is given for Y_R in terms of these a_i. The error involved in the substitution of Y_R for Y in Eq. (1) is evaluated in terms of the anisotropy factor $A=2c_{44}/(c_{11}-c_{12})$ of the crystal, so that a suitable correction can be made.

The foregoing procedure is subject to improvement in two particulars. First, the considerable labor incident to the separate computation of the a_i can be eliminated, with no increase in the complexity of the resultant expression for Y_R. Second, the correction to Y_R in fact depends upon the values of C as well as of A. This circumstance is ignored in reference 1.

The usefulness of the method is further increased by extending the curves for the correction of Y_R to values of A less than unity.

[1] S. L. Quimby and P. M. Sutton, Phys. Rev. **91**, 1122 (1953).
[2] L. Hopf and G. Lechner, Verhandl. deut. physik. Ges. **16**, 643 (1914).

EXPLICIT FORM FOR Y_R

If the expressions for the a_i (reference 1) are substituted in the expression for Y_R and the result simplified, it is found that

$$Y_R=3m_0+\frac{1}{1155}\Big\{[12(15-2x)m_1+288m_2+1112.4m_3+825.6m_4+135m_5]+\frac{1}{c}[54(m_5-m_3)+576(m_4-m_2)]+\frac{54}{d}[m_1-m_3]+\frac{1036.8}{f}[m_3-m_4]+\frac{864}{g}[m_2-m_5]\Big\}, \qquad (3)$$

where

$$\begin{aligned} m_0&=f(0), & & \\ m_1&=f(1)-m_0, & c&=1-2x, \\ m_2&=f(x)-m_0, & d&=3-2x, \\ m_3&=f(2x/3)-m_0, & f&=3-5x, \\ m_4&=f(1-x)-m_0, & g&=3-7x, \\ m_5&=f(1-4x/3)-m_0, & x&=(1-K)/2. \end{aligned}$$

Table I exhibits the variation of Y_R with x and C. The indicated range of values of x covers all the cubic crystals listed by Hearmon[3] with the exception of Na, K, and $NaClO_3$. C, for the cubic crystals listed by Hearmon,[3] ranges from 0.095 to 1.15 except for Na, K, and beta brass.

A correction to Eq. (3) is shown in Figs. 1 (a) and 1 (b). The method of deriving this correction is discussed in a later section.

The computation of a Debye temperature proceeds directly from the elastic constants by calculating C, K, and x. Equation (3) then yields Y_R. This value of Y_R may then be checked by comparing it with Table I. (Equation (3) must be used since interpolation using

[3] R. F. S. Hearmon, Revs. Modern Phys. **18**, 428 (1946).

Table I is of limited accuracy.) This value of Y_R is then corrected using Fig. 1. The estimation of position of curves of C intermediate to those shown rarely will introduce error over $\pm 0.3\%$ in final Y. The curves themselves may be in error by a like amount, so that the error in the computed mean Θ_D, stemming from the computation, may be about 0.2%.

TAYLOR'S SERIES EXPANSION OF Y

As an aid to the evaluation of the effect of C on the correction to Y_R, it is desired to expand Y in ascending powers of x. If $f(z)$ is represented by a polynomial of the nth degree, there are $n+1$ coefficients a_i to determine. Further, if, instead of forcing agreement with $f(z)$ at $n+1$ *different* points, it is required that

$$a_j = f^j(0)/j!,\ \sum_{i=j}^{n} \frac{i!a_i}{(i-j)!} = f^j(1),\ j=0,1,2\cdots m, \quad (4)$$

where $(m+1) \leq (n+1)/2$ and $f^j(z)$ denotes the jth derivative of $f(z)$, then an equation for Y results whose first $m+1$ terms in powers of x are identical with the first $m+1$ terms of the Taylor expansion for Y (see Appendix A).

The series, denoted Y_T, obtained in this manner is

$$\begin{aligned} Y_T = {} & [2C^{-1}+(C+1)^{-1}]-[(6/5)B]x \\ & +[24B+150(C+1)^{-7/2}+105C^{-7/2}]\frac{x^2}{105} \\ & -[-4176B+6000(C+1)^{-7/2}+4440C^{-7/2} \\ & -25620(C+1)^{-9/2}+13335C^{-9/2}]\frac{x^3}{15015} \\ & +[87072B-127200(C+1)^{-7/2}-90480C^{-7/2} \\ & -146160(C+1)^{-9/2}+81900C^{-9/2}+523530(C+1)^{-11/2} \\ & +206246.25C^{-11/2}]\frac{x^4}{255255}\cdots, \quad (5) \end{aligned}$$

where $B=C^{-5/2}-(C+1)^{-5/2}$. The first term is the term for an isotropic body. The second term is that given by Born and v. Kármán.[4,5] The first *three* terms can be obtained from an expansion of Y_R, Eq. (3), since this is the case of $m=2$, $n=5$.

The Taylor's series must diverge for $x \geq |C|$, and converges slowly for $x<|C|$.

Y_T is always less than the true Y for $x<0$ and thus constitutes a lower bound for Y. For $x>0$, the terms of the series alternate in sign so that Y_T alternately bounds above and below as terms are added.

An upper bound for Y can be had for $x<0$ by using $n=5$ and $m=1$ in Eq. (4) and then passing the $[f(z)]_R$ through $z=x$ and $z=1-4x/3$. This $[f(z)]_R$ is then tangent to $f(z)$ at $z=0$ and $z=1$ and lies wholly above $f(z)$ throughout the root ranges. (It can be shown that $[f(z)]_R$ cuts $f(z)$ only at the points chosen.)

TABLE I. Values of Y_R computed by using Eq. (3).

x \ C	0.25	0.50	0.75	1.0
−0.8				5.974
−0.7			21.887	4.614
−0.6			8.916	3.872
−0.5			6.515	3.391
−0.4		15.651	5.332	3.052
−0.3		10.444	4.605	2.802
−0.2	41.231	8.240	4.114	2.613
−0.1	22.508	6.990	3.765	2.466
0	16.716	6.201	3.511	2.354
+0.1	13.968	5.677	3.325	2.267
+0.2	12.461	5.324	3.191	2.203
+0.3	11.608	5.093	3.099	2.158
+0.4	11.128	4.959	3.058	2.131

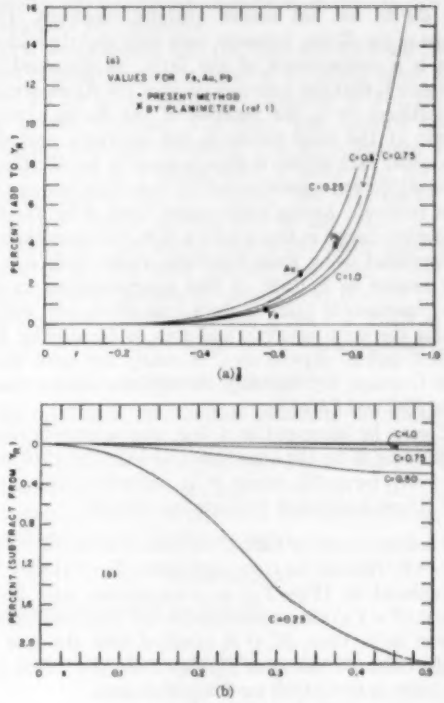


FIG. 1. Correction function E.

ESTIMATE OF CORRECTIONS

The true correction to Y_R is given by

$$E=\sum_i \int_{s_i} \{f(z_i)-[f(z_i)]_R\}\frac{d\Omega}{dz_i}\frac{dz_i}{4\pi}. \quad (6)$$

[4] M. Born, *Atomtheorie des festen Zustandes, Enc. Math. Wiss.* (Teubner, Leipzig), Vol. 3, p. 648.

[5] M. Born and Th. v. Kármán, Physik. Z. 14, 15 (1913). They appear to have neglected to cancel a factor of 4π. See reference 10, page 18.

The $d\Omega/dz_i$ are not known analytic functions. The shapes of the $d\Omega/dz_i$, however, vary only slightly with x. This is a consequence of the facts, as discussed in reference 1, that the intervals (in z) of the $d\Omega/dz_i$ are all proportional to x, the maxima of the $d\Omega/dz_i$ always remain at the same points in the intervals, and the area under each $d\Omega/dz_i$ is always equal to $4\pi/48$. Thus, if the $d\Omega/dz_i$ are approximated by functions having the same intervals, having areas under them of $4\pi/48$ and having no change in shape with x, then the approximate E computed using these functions varies with x in a way similar to E itself. A first approximation to E, using *constants* in place of the actual $d\Omega/dz_i$ and maintaining the same intervals and area, is denoted by E'. Both E and E' depend on C in nearly the same way, since C enters Eq. (6) only through the known functions $f(z_i)$ and $[f(z_i)]_R$.

E' may be improved as a first approximation to E by dividing E' by the quantities $Q(1+Gx)$ for $x>0$ and $Q(1+Hr)$ for $x<0$, where r is defined as the ratio x/C. These factors are evaluated as follows:

Q is determined by taking the ratio of E' to $(Y-Y_R)$ as $x\to 0$. This can be computed since $(Y-Y_R)$ may be be replaced by (Y_T-Y_R) as x approaches zero. Both E' and $(Y-Y_R)$ are proportional to the third and higher powers in x. Thus, E'/Q is *identical* with the true E to the order x^3. Values of Q range from $Q=1.4902$ for $C=1.136$ to $Q=1.5298$ for C equal to zero.

Since E and E'/Q have no lower powers of x than x^3, both go through zero with no slope or curvature. Division by Q serves to adjust E' to agree with E in general magnitude. Since E' is computed with a $d\Omega/dz_i$ of substantially different shape than the true $d\Omega/dz_i$, the size of Q is not unexpected.

For $x>0$, the *true value of* Y, and thus E, can be computed at $x=0.5$ (i.e., at $K=0$) since the three roots of the cubic are here simply α^2, β^2 and γ^2. Division of E'/Q by $(1+Gx)$ further adjusts our error estimate so that proper choice of G gives a good fit to the exact errors at $x=0.5$. Thus, the final error curves in Fig. 1(b) are $E'/Q(1+Gx)$. These curves are precise at either end and have the required lack of curvature at $x=0$. Further, as before stated, E and E' should vary with x in a similar manner, so we expect the shape of these curves to be nearly correct. Judging by the size of the adjustment necessary at $x=0.5$ the largest probable error to be expected in the curves shown is about $\pm 0.3\%$ in Y_R.

As a check on the foregoing method of arriving at the error curves in Fig. 1(b), Y has been calculated by the numerical method of Grüneisen and Goens[6] at the point $C=0.25$ and $x=0.3$. This yields $Y=11.409$ requiring a correction for Y_R of -1.71%. This differs from the Fig. 1(b) value of -1.44% by less than 0.3%. The error in the Grüneisen and Goens numerical method should be an order of magnitude less than the change induced in Y by omitting one of the points in their basic triangle. Omission of the worst point increases Y by only 1.9 percent, so it is felt that the $Y=11.409$ above is in error at worst by 0.5%. Thus, use of Eq. (3) and Fig. 1(b) is of comparable accuracy to the Grüneisen and Goens method—and presents much less computational difficulty.

For $x<0$, division of E' by the function $Q(1+Hr)$ gives the curves plotted in Fig. 1(a). The value of H is obtained by choosing it to give the best fit with values computed graphically for Fe, Au, and Pb.[1] The values of C for Fe, Au, and Pb are, respectively, 1.136, 0.3040, and 0.4248. The resultant curves go through $r=0$ with the required lack of curvature and approach infinity at $r=-1.0$ as they must. Also, as before stated, their shapes should be nearly correct since E and E' should vary with x in a like manner. The fact that *one* parameter, H, is sufficient to give good fit at the three different values of r for Fe, Au, and Pb seems to substantiate the accuracy of the shape of these curves (especially since these three metals cover the useful region of the curves where the error is large and important). Judging by the estimated error in the graphically determined values and the fit obtained, it appears that an estimate of 0.3% error in the curves of Fig. 1(a) is reasonable for the range of r from zero to -0.8.

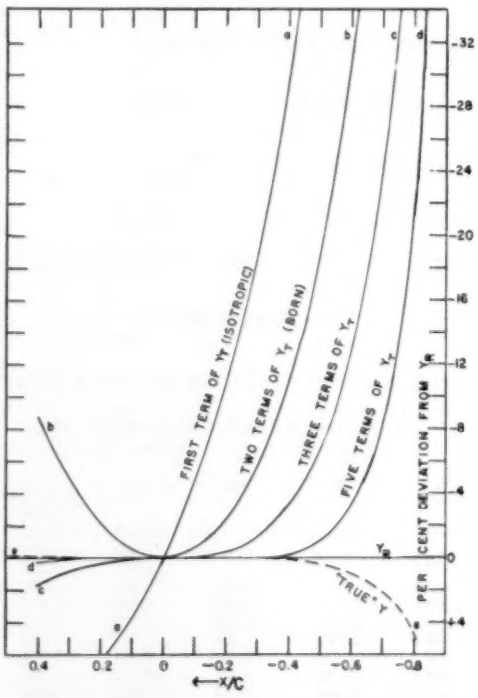


FIG. 2. Comparison of terms of the Taylor series for Y with Y_R, Eq. (3), and the "true" Y. Born's equation is shown to be of low accuracy.

[6] E. Grüneisen and E. Goens, Z. Physik **26**, 255 (1924).

It should be noted that there is good agreement with the Grüneisen and Goens[6] method in this range, too, as mentioned in reference 1. The methods agree to about 0.4% in Y for Au and Ag, *using the same data.* (The trapezoidal averaging method of Grüneisen and Goens can be improved by dividing by the sum of the actual area factors used, rather than using the factor π/n. This raises Y values usually about 0.6%. The need for this improvement is easily demonstrated by applying their method to the isotropic case.)

In summary, it is estimated that the curves of Fig. 1 are accurate to about 0.3%. Interpolation for intermediate C values may double this error. The percent error in the computed Θ_D introduced by the substitution of the corrected Y_R for Y in Eq. (1) is one-third the percent difference between them. Therefore, the use of Eq. (3) and Fig. 1 should result in the mean Θ_D accurate to about 0.1% on a given curve or, when interpolating curves, to about 0.2%. This is about an order of magnitude less than the errors introduced in Θ_D by the present experimental uncertainty in the c_{ij}.

Use of Eq. (3) and Fig. 1 reduces the arrival at an accurate Θ_D to a short operation taking, at most, a few minutes. This makes possible computations of Θ_D for many values of elastic data without being prohibitive in time and effort, and enables the experimenter to compute easily the effect on Θ_D of variation of such parameters as temperature, cold work, radiation damage, or alloy proportion.

COMPARISON OF FORMULAS

Figure 2 compares terms of Y_T to Y_R. The dashed curve is the "true" Y as discussed above. It is apparent that Y_R fits the "true" curve best. For positive r, Y_T to the term in x^4 is nearly as good as Y_R; for negative r, it fails much beyond $r=-0.5$. It appears that Y_T would have to be carried out several further terms in order to compete with Y_R near $r=-0.7$ where many of the metals occur. (As a consequence, Y_T is chiefly of value in computing Q.) Also, the equation given by Born[4,5] is quite inaccurate except close to isotropy, say $r<|0.1|$. (From Hearmon's Table,[3] only tungsten falls in this category.) Figure 2 may be used as a guide to use of Y_T for substances of small x. It is plotted for $C=1$ but the variation of C does not change these curves substantially.

Two papers by Blackman seem to complement the method here described. The first permits calculation of Θ_D for cases where $(c_{12}+c_{44})\ll c_{11}$, i.e., for very positive r.[7] The second applies where $c_{11}-c_{12}$ is small, i.e., $r\doteq -1$.[8] The formulas he proposes are relatively simple but may be in error by several percent.

Bhatia and Tauber[9] give an equation for Y which is also accurate to several percent. This is derived using the first three Kubic Harmonics for the integrand. Their equation contains the same m_i as Eq. (3), but the coefficients are constants.

It is worth noting that the Hopf and Lechner[2] treatment necessarily gives a value of Θ_D which is too high (low Y) for all substances of negative x. The metals are in this category. The Hopf and Lechner representative function intersects $f(z)$ only between $z=0$ and $z=1$ and as a consequence is lower than the true $f(z)$ throughout the root ranges. The wide use of the Hopf and Lechner method may be one reason for the fact that Θ_D from elastic data is frequently high compared to the Θ_D from specific heat data.[10,11] As an example, even for nearly isotropic Al ($r=-0.1561$)[12] the Hopf and Lechner method gives a Y which is 1.7 percent low; for more negative r it worsens rapidly.

ACKNOWLEDGMENTS

In conclusion, the author gratefully acknowledges the helpful counsel and encouragement of Professor S. L. Quimby of Columbia University. The author is indebted to the Corning Glass Works and to Dr. W. H. Armistead, Director of Research and Development, for permission to publish this material.

APPENDIX A

Equation (1) may be written

$$Y=\frac{1}{4\pi}\sum_i\int_{z_i} f(z_i)\frac{d\Omega}{dz_i}dz_i, \tag{1A}$$

where $(d\Omega/dz_i)dz_i$ is the solid angle associated with the ith root lying between z_i and z_i+dz_i. The defining equations for y_i are

$$\begin{aligned} z_1&=1+y_1x,\\ z_2&=y_2x,\\ z_3&=y_3x. \end{aligned} \tag{2A}$$

Then Eq. (1A) becomes

$$Y=\frac{1}{4\pi}\sum_i\int_0^{k_i} f(z_i)\left(x\frac{d\Omega}{dz_i}\right)dy_i, \tag{3A}$$

where the k_i are appropriate *constants* (all ranges of z_i are proportional to x). Also

$$\int_0^{k_i}\left(x\frac{d\Omega}{dz_i}\right)dy_i=\int_{z_i}\frac{d\Omega}{dz_i}dz_i=4\pi. \tag{4A}$$

Taylor's series for $f(z)$ at $z=0$ for z_2 and z_3 and at $z=1$ for z_1 yields ($|k_2|>|k_3|$ and $y_2=y_3=y$)

$$Y=\frac{1}{4\pi}\left[\int_0^{k_1}\left(x\frac{d\Omega}{dz_1}\right)\sum_j x^jy_1^j\frac{f^{(j)}(1)}{j!}dy_1 + \int_0^{k_2}\left[x\frac{d\Omega}{dz_2}+x\frac{d\Omega}{dz_3}\right]\sum_j x^jy^j\frac{f^{(j)}(0)}{j!}dy\right]. \tag{5A}$$

[6] M. Blackman, Proc. Roy. Soc. (London) **A149**, 126 (1934).
[7] M. Blackman, Phil. Mag. **42**, 1441 (1951).
[8] A. B. Bhatia and G. E. Tauber, Phil. Mag. **45**, 1211 (1954).
[9] M. Blackman, Repts. Progr. Phys. **8**, 19 (1941).
[10] F. Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), p. 111.
[11] P. M. Sutton, Phys. Rev. **91**, 816 (1953), data at 0°K.

On the other hand, use of a representative polynomial

$$[f(z)]_R=\sum_{j=0}^{n} a_j z^j \tag{6A}$$

yields

$$Y=\frac{1}{4\pi}\left[\int_0^{h_1}\left(x\frac{d\Omega}{dz_1}\right)\sum_{j=0}^{n}\left(x^j y_1{}^j\sum_{i=j}^{n}\frac{i!a_i}{(i-j)!j!}\right)dy_1 + \int_0^{h_2}\left(x\frac{d\Omega}{dz_2}+z\frac{d\Omega}{dz_2}\right)\sum_{j=0}^{n}x^j y^j a_j dy\right]. \tag{7A}$$

A jth term in Eq. (5A) or (7A) must be at least of order j in powers of x since it is proportional to

$$\left|x^j\int_0^{k_i}\left(x\frac{d\Omega}{dz_i}\right)y^j dy\right|\leq\left|x^j k_i{}^j\int_0^{k_i}\left(x\frac{d\Omega}{dz_i}\right)dy\right|=|4\pi x^j k_i{}^j|,$$

and the inequality would fail for x sufficiently near zero if the jth term were of order less than j in x.

Therefore, the requirement of Eq. (4) forces Y_R to be *identical* to Y for all powers of x^j where $j\leq m$ as can be seen by comparing Eq. (5A) with Eq. (7A). Equation (4) constitutes $2(m+1)$ conditions on the $(n+1)$ coefficients a_i so that $(n+1)\geq 2(m+1)$. Thus expansion of Y_R in powers of x yields the first $(m+1)$ terms in the Taylor's series for Y in powers of x.

The Y_R needed to give the Y_T of Eq. (5) has $n=9$ and $m=4$. Hopf and Lechner[2] give $\sum z_i{}^j$ up to $j=5$ (case of $n=5$). To obtain Y_R for $n>5$, use is made of the recurrence relation

$$\sum z_i{}^j=\sum z_i{}^{j-1}-K_1\Gamma\sum z_i{}^{j-2}+K_2\chi\sum z_i{}^{j-3}, \tag{8A}$$

where

$$K_1=(1-K^2),$$
$$K_2=(1-3K^2+2K^3),$$
$$\Gamma=\alpha^2\beta^2+\beta^2\gamma^2+\gamma^2\alpha^2,$$
$$\chi=\alpha^2\beta^2\gamma^2.$$

Equation (8A) is obtained directly from the basic cubic in z. The eleven integrals needed for $n=9$ may then be obtained, and give the following averages over the unit sphere: Γ, 1/5; χ, 1/105; Γ^2, 1/21; $\Gamma\chi$, 1/385; Γ^3, 61/5005; χ^2, 1/5005; $\Gamma^2\chi$, 1/1365; Γ^4, 277/85085; $\Gamma\chi^2$, 5/85085; $\Gamma^3\chi$, 205/969969; χ^3, 5/969969. Hopf and Lechner[2] give the first four of these.

The final equation for Y_R for $n=9$ is[12]

$$\begin{aligned}Y_R=3a_0&+\sum_{j=1}^{9}a_j-\frac{K_1}{5}\sum_{j=2}^{9}ja_j+\frac{K_2}{105}\sum_{j=3}^{9}ja_j\\&+\frac{K_1^2}{21}\sum_{j=4}^{9}\frac{j(j-3)}{2}a_j-\frac{K_1K_2}{385}\sum_{j=5}^{9}j(j-4)a_j\\&-\frac{61K_1^3}{5005}\sum_{j=6}^{9}\left(\frac{j}{3}\sum_{n=5}^{j}(j-n)\right)a_j+\frac{K_2^2}{5005}\sum_{j=6}^{9}\frac{j(j-5)}{2}a_j\\&+\frac{K_1^2K_2}{1365}\sum_{j=7}^{9}\left(j\sum_{n=6}^{j}(j-n)\right)a_j+\frac{277K_1^4}{85085}(2a_8+9a_9)\\&-\frac{5K_1K_2^2}{85085}(8a_8+27a_9)-\frac{205(K_1^3K_2)}{969969}(9a_9)\\&+\frac{5K_2^3}{969969}(3a_9).\end{aligned} \tag{9A}$$

[12] Terms of Eq. (9A), without a_9, are given by K. Fuchs in powers of K in Proc. Roy. Soc. (London) **A153**, 622 (1936) and **A157**, 444 (1936). An error in the coefficient of a_8 in **A153** is corrected in **A157**.

Photoelectric Emission from Polycrystalline Graphite

E. Taft and L. Apker
General Electric Research Laboratory, Schenectady, New York
(Received May 26, 1955)

Energy distribution of the photoelectrons from polycrystalline graphite were measured by retarding-potential methods in spherical phototubes with small, central, interchangeable emitters. There were distinct deviations from metallic behavior. The results were consistent with a density of states that was relatively small at the Fermi level and that rose to a value several-fold higher 1 ev away. The photoelectric yield from graphite is about ten times below that of common metals.

I. INTRODUCTION

IN the course of other work, we have had occasion to use polycrystalline graphite as a photoemitter. A rather wide variety of samples was inspected. Often, no rigorous measures were taken to clean or degas the samples. Nevertheless, the photoelectric behavior of these graphites has been surprisingly uniform in character. Hence, one suspects that the results may be more significantly influenced by the electronic energy structure of the material than by accidental circumstances of measurement. With this possibility in mind, we present some typical results in this paper.

II. EXPERIMENTAL DETAILS

Other papers describe the methods of investigation.[1] Some samples consisted of aquadag painted on quartz and baked at about 500°C in vacuum. Others were bulk graphite, machined to shape and degassed by induction heating at temperatures up to 2000°C. In two tubes, carbon was evaporated from graphite spirals onto metal or quartz substrates.

III. RESULTS

Energy distributions were obtained by differentiating current-voltage characteristics for small, central, graphite emitters in the usual type of phototube with a large spherical collector. The 0°K stopping potential, V_0, was determined from data on interchangeable metallic emitters. Figure 1 shows a typical set of results. It is clear that graphite deviates measurably from a simple metal in its behavior, as shown explicitly by the curves for $h\nu = 5.78$ ev. There is a sparsity of higher energy photoelectrons. The results are characteristic of a low density of states in the vicinity of the Fermi level, a condition especially susceptible to detection by the present methods. At a point roughly 1 ev below the Fermi level, the density of occupied electron energy levels is noticeably higher (see Fig. 2).

This type of behavior is what one would expect from well-known theoretical attacks on the graphite problem.[2] All we can claim here is that we have observed this behavior on all the graphite samples we have measured. To this extent our results support the theoretical conclusions.

In principle, it is conceivable that by this method one could look for the predicted minimum in the density of states at or very near the Fermi level. We prefer not to do this with our present results because the process becomes untrustworthy more than 0.1 to 0.2 ev above the Fermi level, where the probability that levels are occupied becomes small. Spurious results may arise from stray radiation of higher photon energy or from slight non-uniformity in the work function of the large col-

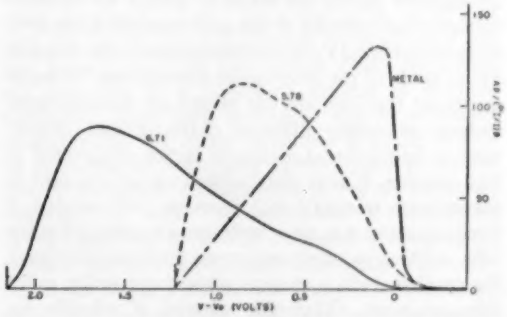


FIG. 1. Typical normalized energy distributions for photoelectrons from graphite emitters. The pertinent values of $h\nu$ are given next to the curve, which were obtained by differentiation of current-voltage characteristics; errors in the ordinates do not exceed 10 percent. The units on the ordinate scale are percentages of saturation current I_{10} (measured at $V = 10$ volts) per volt. On the abscissa scale, kinetic energies of external photoelectrons increase toward the right; the zero point corresponds to photoelectrons originating at the Fermi level in the emitter. For comparison, a curve is given for a typical simple metal with the same work function $\varphi = 4.56$ ev as the graphite and with $h\nu = 5.78$ ev. Note the relative sparsity of higher energy photoelectrons from the graphite as compared with the metal.

[1] Apker, Taft, and Dickey, Phys. Rev. **74**, 1462 (1948); **84**, 508 (1951); E. Taft and L. Apker, Phys. Rev. **96**, 1496 (1954); H. B. Huntington and L. Apker, Phys. Rev. **89**, 352 (1953); H. B. Huntington, Phys. Rev. **89**, 357 (1953).

[2] P. R. Wallace, Phys. Rev. **71**, 622 (1947), finds a minimum in the density of electron energy states at the Fermi level, where the valence and conduction bands are just contiguous. In our work, an additional effect may contribute to the rise in efficiency of photoelectric excitation for states below the Fermi level: the transition probability rises linearly with energy (i.e., with $V - V_0$ in Fig. 1) as one departs from the edge of the valence band. For more recent progress, see J. L. Carter and J. A. Krumhansl, J. Chem. Phys. **21**, 2238 (1953); J. E. Hove, Phys. Rev. **97**, 1717 (1955); W. M. Lomer, Proc. Roy. Soc. (London) **A227**, 330 (1955); D. F. Johnston, Proc. Roy. Soc. (London) **A227**, 349 (1955); D. F. Johnston, Proc. Roy. Soc. (London) **A227**, 359 (1955). Results consistent with these have been seen in soft x-ray work by H. W. B. Skinner [Phil. Trans. **A239**, 95 (1940)] and by F. C. Chalklin [Proc. Roy. Soc. (London) **194**, 42 (1948)].

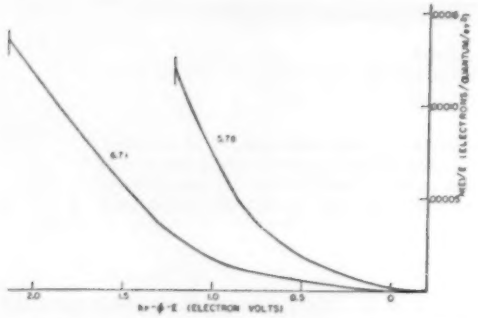


FIG. 2. $N(E)/E$ vs $h\nu-\phi-E$ for the same emitter represented in Fig. 1. As discussed in reference 1, this function is crudely similar to the product of occupied state density and photoelectric transition probability. The important point here is that the "density of occupied states" at the Fermi level, $h\nu-\phi-E=0$, is small relative to that one volt away.

lector. Such effects tend to make the Fermi "tail" too high, in a manner that would be difficult to distinguish from an increasing density of electron energy states. Below the Fermi level, these difficulties are not serious.

Patch fields due to the polycrystalline character of the emitters should not seriously distort the current-voltage characteristics in the neighborhood of the 0°K stopping potential V_0, which corresponds to the common Fermi levels of the interchangeable emitters. It seems significant that reproducible results are obtained from surfaces of widely different character and origins, ranging from baked aquadag to carbon evaporated in high vacuum. Energy distributions are not changed in nature when relatively high pressures (~1 micron) of foreign gases such as air or hydrogen are admitted to the tube and then pumped away; even the work function of the graphite surface remains surprisingly stable under this treatment. (Thus the material is valuable for electrode surfaces such as the spherical collectors in these phototubes).

The photoelectric yield from these emitters is shown in Fig. 3. It was of order 10^{-5} electron per incident photon at $h\nu=5.78$ ev, about 10 times lower than the value for common metals of the same work function. (Again, this characteristic is valuable in electrodes,

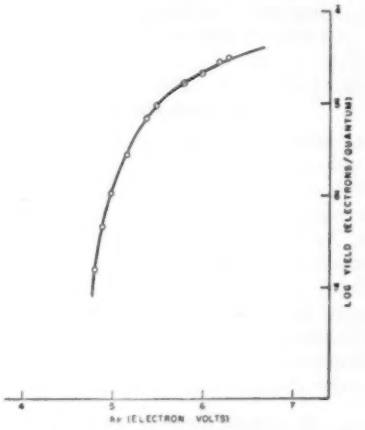


FIG. 3. Logarithm of the yield at room temperature per incident photon. The yield several volts beyond the work function, 4.56 ev, is about ten times below that for common metals.

since reverse currents are suppressed.) Work functions for the seven samples measured in detail were 4.48, 4.51, 4.56, 4.70, 4.73, 4.80, and 4.88 ev. These are in reasonable agreement with the work functions of the aquadag collectors in our phototubes, more than 90% of which lie between 4.65 and 4.85 ev.[3]

[3] See reference 1, first paper, footnote 29. Aquadag-bentonite mixtures, which we have also used frequently, have higher work functions lying between 5.2 and 5.5 ev.

Multiple Electron Excitation in Auger Processes*

MAX WOLFSBERG AND MORRIS L. PERLMAN
Chemistry Department, Brookhaven National Laboratory, Upton, New York

(Received April 18, 1955)

Consideration of the Auger effect as a polyelectron phenomenon leads to the expectation of an accumulation of a greater charge on an atom, as the result of an Auger cascade, than that obtainable by the conventional two-electron theory. On this basis, calculations are made of the probability for the accumulation of large charges by chlorine atoms originally possessing a K vacancy. The theoretical results correlate well with the recent experimental measurements of the charge distribution of the chlorine atoms produced by the electron-capture decay of argon-37.

THE Auger process is usually described as a process involving two bound atomic electrons, one of which undergoes a transition to a vacant orbital of lower energy while the other one is ejected to the continuum. The charge on the atom is thus increased by one unit.

In the nonrelativistic theory of the Auger effect, the probability of a transition between two atomic states of the same energy ψ_i and ψ_f is proportional to M,[1] where

$$M=\left|\int\psi_f^*\sum_{s>t}\frac{e^2}{r_{st}}\psi_i d\tau\right|^2. \quad (1)$$

In the evaluation of M, ψ_f and ψ_i are usually taken to be properly antisymmetrized two-electron atomic orbital wave functions for the two electrons involved in the process; the other electrons in the polyelectronic atom are ignored. In the atomic orbital approximation, this procedure is justified if the assumption is made that the orbitals of the other electrons are the same in both ψ_f and ψ_i and thus have orthogonality properties such that e^2/r_{st} can connect only states which differ in the orbitals of two electrons.

Since one generally finds that the orbitals of the other electrons also must change due to a change in screening constants, the above description of the process as a two-electron problem is not quite complete. Those bound electrons which have velocities small compared to that of the ejected "Auger" electron suffer a "sudden perturbation" due to the sudden change in effective nuclear charge and may undergo a transition to an excited state. This situation is similar to the case of the sudden perturbation in beta decay, which has been studied by a number of investigators.[2] Thus for these electrons, the wave function ψ_f in Eq. (1) should be written with effective values of the nuclear charge computed with allowance for the missing Auger electron. There exists then a probability that an atom may gain more than one charge as a result of an Auger process, if one of the other electrons is excited either to a continuum state or to a bound state capable of participating later in an Auger process which would otherwise not have been possible.

Computations of the above effect have been made for Auger transitions in chlorine, which are of special interest on account of recent experimental measurements of the average charge[3] and also of the actual charge distribution[4] of chlorine atoms resulting from the electron capture decay of argon-37. The effective Z values for the various atomic orbitals were found using the Slater recipes.[5] From the values of ΔZ and with the aid of tabulations made by Schwartz,[2] one sees immediately that only certain Auger transitions will give rise to a significant amount of multiple excitation.. These cases include $3s$ and $3p$ excitation accompanying the Auger transition $K\rightarrow LL$;[6] $3s$ and $3p$ excitation accompanying $L_{\rm I}\rightarrow L_{\rm II,\,III}M$; $2s$, $2p$, $3s$, and $3p$ excitation accompanying $L\rightarrow MM$. Of these, however, $3s$ and $3p$ excitation accompanying $L_{\rm I}\rightarrow L_{\rm II,\,III}M$ and $2s$ and $2p$ excitation accompanying $L\rightarrow MM$ are both energetically impossible. The other two above Auger processes which may be accompanied by $3s$ and $3p$ excitation fulfill not only the energy requirements but also the requirements which make the sudden perturbation treatment applicable. By means of the Auger phenomenon alone chlorine of charge plus 5 can be produced following the K-capture decay of argon by the following set of five processes: $K\rightarrow L_{\rm I}L_{\rm I}$, $L_{\rm I}\rightarrow L_{\rm II,\,III}M$, $L_{\rm I}\rightarrow L_{\rm II,\,III}M$, $L_{\rm II,\,III}\rightarrow MM$, $L_{\rm II,\,III}\rightarrow MM$. The same charge may be produced also by a smaller number of Auger processes accompanied by sudden perturbation excitation. The principal amount of $3s$ and $3p$ excitation accompanies the $K\rightarrow L_{\rm I}L_{\rm I}$ process; a very minor amount accompanies $L_{\rm II,\,III}\rightarrow MM$. Substituting in Eq. (1) one obtains for M, with no sudden perturbation excitation, for the $K\rightarrow L_{\rm I}L_{\rm I}$

* Research carried out under the auspices of the U. S. Atomic Energy Commission.

[1] E. H. S. Burhop, *The Auger Effect* (Cambridge University Press, London, 1952).

[2] E. L. Feinberg, J. Phys. (U.S.S.R.) **4**, 424 (1941); H. M. Schwartz, J. Chem. Phys. **21**, 45 (1953); H. Primakoff and F. T. Porter, Phys. Rev. **89**, 930 (1953).

[3] M. L. Perlman and J. Miskel, Phys. Rev. **91**, 182 (1954); S. Wexler and T. H. Davies, Phys. Rev. **93**, 182 (1954).

[4] O. Kofoed-Hansen, Phys. Rev. **96**, 1045 (1954); A. H. Snell and F. Pleasanton, Phys. Rev. **98**, 1174(A) (1955).

[5] J. C. Slater, Phys. Rev. **36**, 57 (1930).

[6] The hole nomenclature is employed here to designate Auger transitions.

transition,

$$M=\Big|\int 1s_f^*(1)\alpha(1)F^*(2)\beta(2)1s_f^*(3)\beta(3)2p_{a\,f}^*(4)\alpha(4)$$
$$\times 2p_{a\,f}^*(5)\beta(5)\cdots\times\sum_{i>j}\frac{e^2}{r_{ij}}\sum_{\nu}(-1)^{\nu}P_{\nu}2s_i(1)\alpha(1)\quad(2)$$
$$\times 2s_i(2)\beta(2)1s_i(3)\beta(3)2p_{a\,i}(4)\alpha(4)2p_{a\,i}(5)\beta(5)\cdots d\tau\Big|^2,$$

where α and β refer to the two types of spin functions, F stands for the wave function of the ejected electron, P_ν represents the permutation of electrons, and ν is even or odd according to whether the permutation is even or odd. The quantity which is squared in Eq. (2) consists, when expanded, of a sum of terms, each of which is a product of a two-electron integral involving e^2/r_{ij} multiplied by a number of one-electron overlap-type integrals. Equation (2) reduces to

$$M=|A|^2K,\qquad(3)$$

with

$$A=\int 1s_f^*(1)\alpha(1)F^*(2)\beta(2)\times\frac{e^2}{r_{12}}2s_i(1)\alpha(1)2s_i(2)\beta(2)d\tau_1d\tau_2$$

and

$$K=\Big|\int 1s_f^*1s_fd\tau\Big(\int 2p_f^*2p_fd\tau\Big)^6\Big(\int 3s_f^*3s_fd\tau\Big)^2$$
$$\times\Big(\int 3p_f^*3p_fd\tau\Big)^6-\Big[6\int 2p_f^*3p_fd\tau\int 3p_f^*2p_fd\tau$$
$$+\int 1s_f^*3s_fd\tau\int 3s_f^*1s_fd\tau\Big]\times(\text{quantity of the order unity})+\text{smaller terms}\Big|^2,$$

If one omits all terms not containing the two electron integral A, which is the e^2/r_{ij} integral corresponding to the Auger transition as it is usually taken. The terms involving other e^2/r_{ij} integrals are considered negligible, although this factor has not been thoroughly explored. The probability, M, of obtaining a final state with sudden perturbation excitation is similar to Eq. (3) with K suitably modified to indicate the orbitals which correspond to excitation in the final state. K, when summed over all possible final states for the electrons not taking part in the $K\rightarrow LL$ Auger transition, should be about equal to unity. This summed probability differs slightly from unity chiefly because transitions to $1s_f\alpha$ and $F\beta$ are not allowed.

Both $\int 1s_f^*1s_fd\tau$ and $\int 2p_f^*2p_fd\tau$ are very close to unity, so that the deviation of K in Eq. (3) from unity gives the probability of the $3s$ and $3p$ electrons not being in their original configuration. If one subtracts from this number the probability that the $3s$ electron makes a transition to the empty $2s$ orbital, one obtains the probability that the $3s$ and $3p$ electrons are actually excited during the Auger process. Integrals of the type appearing in Eq. (3) have usually been evaluated with hydrogenic atomic orbitals.[2] One finds, using Slater Z values and hydrogenic functions, that the probability of exciting $3s$ and $3p$ electrons in the foregoing Auger cascade leading to the plus five charged chlorine is 0.16; 98 percent of this total arises from the $K\rightarrow L_IL_I$ transition. Hydrogenic atomic orbitals are constructed in such a way that a $3s$ function corresponding to a given Z is orthogonal to a $2s$ function with the same value of Z. In the atom, the Z of the $2s$ function is of course different from that of the $3s$ function. The probability of excitation has therefore also been evaluated using Slater orbitals which are actually orthogonalized. The result obtained in this way for the excitation of $3s$ and $3p$ electrons is 0.12. While this number is close to that obtained with the hydrogenic functions, it should be noted that the bracketed quantity in the expression for K is negligible in the case of the orthogonal functions but is quite sizeable in the case of the hydrogenic functions.

Using hydrogenic functions, one finds that the probability of actually ionizing $3s$ and $3p$ electrons is only 0.03. However, some of the integrals computed here and calculations made by Levinger[7] indicate that the ionization probability may be much higher than that computed with hydrogenic wave functions and, correspondingly, that the probability of excitation to bound states may be much lower. Moreover, the excited electrons may also take part in Coster-Kronig transitions of the type $M_I\rightarrow M_{II,III}X$, where X corresponds to the orbital to which the $3s$ or $3p$ electrons had been excited. How probable this Coster-Kronig transition may be compared with other processes which compete to fill the M_I hole or deexcite the X electron, one cannot easily say. One does know, however, that for outer shell orbitals radiative transitions are much less probable than Auger processes. Thus, under proper conditions, every excited $3s$ and $3p$ electron may be ionized, the result being the accumulation of a positive charge on the chlorine atom larger than five units, the maximum which may be produced by means of the usual Auger mechanisms following the creation of a K vacancy. Recent measurements by Snell and Pleasanton[4] of the chlorine atom charge distribution show that there is 10 percent probability for Cl^{+6}, 1.8 percent for Cl^{+6}, and 0.4 percent for Cl^{+7}. Snell and Pleasanton's results indicate that some sudden perturbation "ionization" accompanies the Auger processes.[8]

[7] J. S. Levinger, Phys. Rev. **90**, 11 (1953).

[8] In a private communication, A. H. Snell has pointed out that these high charges may be produced also by the occurrence of Auger transitions $M_I\rightarrow M_{II,III}M_{II,III}$. It cannot be stated definitely that such transitions are energetically possible following the decay of A^{37}; however, from the compilation of ionization

If one analyzes their results, assuming that the probability of sudden perturbation "ionization" is the same (except for a statistical factor) for all $K \rightarrow LL$ Auger processes (the main Auger processes for exciting $3s$ and $3p$ electrons), one finds that they are consistent with a probability of about 0.20 for exciting the $(3s)^2(3p)^6$ shell in the Auger cascade.[9] This probability is of the same order of magnitude as the one calculated above from theory. The theoretical value could change considerably, however, if more exact wave functions were available.

ACKNOWLEDGMENTS

The authors are greatly indebted to Dr. William Rubinson and to Dr. A. H. Snell for illuminating discussions on the subject of radiationless transitions.

energies by Hill, Church, and Mihelich [Rev. Sci. Instr. **23**, 523 1952)], the indication is that these transitions may be exoergic.

[9] This calculation was made on the basis that the probability of ionizing two electrons as a result of the sudden perturbation effect is just the square of the probability of ionizing one electron. This may be an oversimplification when the ionization proceeds by the mechanism $M_{\mathrm{I}} \rightarrow M_{\mathrm{II,III}}X$, since then the production of the plus seven charge requires a more severe restriction on the sequence of processes than the production of the plus six charge.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Correlation Energy of Helium

HENRI MITLER
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received March 25, 1955)

The correlation energy of helium is calculated by using an approximate central potential, in a manner analogous to Slater's calculation of the exchange energy.

IN the Hartree self-consistent field equations for an n-electron atom of atomic number Z,[1]

$$\mathcal{H}_i u_i \equiv -\nabla_i^2 u_i - 2Z\frac{u_i}{r_i} + 2\left[\sum{}' \int \frac{|u_k|^2}{r_{ik}} d\tau_k\right] u_i = \epsilon_i u_i, \quad (1)$$

(where atomic units have been used: $e^2=2$, $\hbar^2/2m=1$), it is well known that the equation can be interpreted as a Schrödinger equation in which the third term on the left side is the interaction energy of each electron with an averaged-out, spherically symmetric charge distribution of all the other electrons. Thus ϵ_i is approximately the dissociation energy of the ith electron. In the sum $\sum \epsilon_i$, the interaction energies are taken into account twice; hence

$$E \equiv \langle H \rangle = \sum_i \epsilon_i - \sum_{i>k} \int\int \frac{|u_k|^2 |u_i|^2}{r_{ik}} d\tau_k d\tau_i, \quad (2)$$

where H is the Hamiltonian of the system, and the expectation value is taken with the Hartree function for the ground state,

$$\psi_H = u_1 u_2 \cdots u_n, \quad (3)$$

found from (1).

For helium, the Hartree-Fock function, which takes the symmetry of the wave function into account, is identical with the Hartree function, and $u_1 = u_2$, $\epsilon_1 = \epsilon_2 \equiv \epsilon$, so that[2] Eq. (2) yields

$$E_{\mathrm{He}} = 2\epsilon - \int\int \frac{|u_1|^2 |u_2|^2}{r_{12}} d\tau_1 d\tau_2. \quad (4)$$

Wilson and Lindsay[3] calculated the Hartree $\psi(1s)$ function and its eigenvalue ϵ for helium, and found

$$\epsilon = -1.836 \text{ Rydbergs (ry)}, \quad E_{\mathrm{He}} = -5.723 \text{ ry}.$$

This value is to be compared with the experimental value,

$$E_{\mathrm{exp}} = -5.807 \text{ ry}.$$

[Note that according to reference 2, we indeed find $\epsilon - 4R_{\mathrm{He}}/R_{\mathrm{H}} = -5.836$ ry.]

A self-consistent field calculation does not take into account, however, the instantaneous interactions among the electrons. These interactions give rise to a *correlation* in their positions (and hence a shift in the energy). For He in the ground state, the electrons have opposite spin so that there is no interaction between them due to the Fermi statistics, and the energy of correlation will arise solely from the Coulomb interaction. As good a definition of correlation energy as any, defines it as the difference between the energy as calculated from the Hartree-Fock procedure and the exact energy.[4]

Wigner[5] has given a free-electron gas approximation to this correlation energy for alkali atoms in solids, in the form of an additional central "correlation" potential that the electrons of $\pm$ spin are subject to, in the

[1] F. Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), p. 677 *ff.*, esp. Eq. (4), p. 677.

[2] Since the Hartree energy parameter $|\epsilon|$ has been assumed to be approximately equal to the first ionization potential, μ, then in the ground state,

$$E_{\text{experimental}} = -\mu - \frac{Z^2 R_{\mathrm{H}}}{n^2 R_{\mathrm{H}}} = -\mu - 4\frac{R_{\mathrm{He}}}{R_{\mathrm{H}}} \sim \epsilon - 4\frac{R_{\mathrm{He}}}{R_{\mathrm{H}}},$$

where the energy is, as before, in Rydbergs (ry).

[3] V. C. Wilson and R. B. Lindsay, Phys. Rev. **47**, 681 (1935).

[4] See, however, Green, Mulder, and Miller, Phys. Rev. **91**, 35 (1953); Green, Lewis, Mulder, Wyeth, and Woll, Phys. Rev. **93**, 273 (1954).

[5] E. Wigner, Phys. Rev. **46**, 1002 (1934); Trans. Faraday Soc. **34**, 678 (1938).

Hartree-Fock equations:

$$V(r)=\frac{0.576e^2}{5.1+r_s(r)}, \tag{5}$$

where

$$|u(r)|^2=\rho_\mp(r)=\tfrac{1}{2}[(4\pi/3)r_s^3]^{-1} \tag{6}$$

is the local density of electrons of the opposite spin. If we assume that this holds for the isolated atom, we can compute a correlation correction quite simply. Such a procedure for the treatment of the correlation energy is entirely analogous to Slater's free-electron gas estimation of the exchange energy.[6]

Taking $v(r)$ as the modified wave function, the modified equation we must solve is

$$\left[-\nabla^2-\frac{4}{r}+2\int\frac{|v_1|^2d\tau_1}{|\mathbf{r}-\mathbf{r}_1|}-V(r)\right]v(r)=\varepsilon v(r). \tag{7}$$

Defining

$$P\equiv(4\pi)^{\frac{1}{2}}ru, \tag{8}$$

and

$$\frac{2Z_P(r)}{r}\equiv\frac{2Z}{r}-2\int\frac{|P(r_1)|^2}{|\mathbf{r}-\mathbf{r}_1|}d\tau_1, \tag{9}$$

Eq. (1) becomes for $Z=2$,

$$P''(r)=-[2Z_P(r)/r+\epsilon]P(r), \tag{10}$$

and similarly the modified equation (7) becomes, with $R\equiv(4\pi)^{\frac{1}{2}}rv$,

$$R''(r)=-\left[\frac{2Z_P(r)}{r}+\varepsilon+\frac{1.152}{5.1+r_s}\right]R(r). \tag{7a}$$

Strictly speaking, if we are to solve (7a) by a self-consistent field method, we should use Z_R rather than Z_P, Z_R being defined in a manner analogous to Z_P in (9) [using $R(r)$ in lieu of $P(r)$]. For the approximations we are making, however, the error will be negligible. For the solution of (7a), Noumerov's method is quick, and sufficiently accurate.[7] We may obtain a good estimate of the correction, however, by using first-order perturbation theory.

We substitute in Eq. (7), with λV instead of V,

$$v=u+\lambda u'+\lambda^2u''+\cdots, \quad \varepsilon=\epsilon+\lambda\epsilon'+\lambda^2\epsilon''+\cdots.$$

Expanding and equating the coefficients of equal powers of λ, we obtain

$$\mathcal{H}u=\epsilon u, \tag{1}$$

$$\mathcal{H}u'+4\int\frac{u_1u_1'}{|\mathbf{r}-\mathbf{r}_1|}d\tau_1u-V(r)u=\epsilon'u+\epsilon u', \tag{11}$$

etc. Multiply both sides of (11) by u and integrate over the whole volume. Then since $\mathcal{H}$ is Hermitian, we obtain

$$\epsilon'=-\langle V(r)\rangle+4\left\langle\int\frac{u_1u_1'd\tau_1}{|\mathbf{r}-\mathbf{r}_1|}\right\rangle. \tag{12}$$

From (4), the perturbed energy will be

$$E'=2(\epsilon+\epsilon')-\left[\int\int\frac{(u_1+u_1')^2(u_2+u_2')^2}{r_{12}}d\tau_1d\tau_2-\int Vu^2d\tau\right]$$

$$=2\epsilon+2\epsilon'-\int\int\frac{u_1^2u_2^2d\tau_1d\tau_2}{r_{12}}-4\left\langle\int\frac{u_1u_1'd\tau_1}{|\mathbf{r}-\mathbf{r}_1|}\right\rangle$$

$$+\langle V\rangle+\text{higher order terms}.$$

Note that one of the ϵ' correction terms for the eigenvalue is canceled by the term $\langle V\rangle$. Hence, from (4) and (12),

$$E'=E-\langle V\rangle+4\left\langle\int\frac{u_1u_1'd\tau_1}{|\mathbf{r}-\mathbf{r}_1|}\right\rangle=E+\epsilon', \tag{13}$$

which was to be expected because of reference 2. The normalization, however,

$$\int(u+u')^2d\tau=1,$$

implies that

$$2\int uu'd\tau=-\int(u')^2d\tau, \tag{14}$$

so that the term

$$4\left\langle\int\frac{u(r_1)u(r_1)'d\tau_1}{|\mathbf{r}-\mathbf{r}_1|}\right\rangle$$

is of the second order and can therefore be dropped. Equation (13) becomes

$$E'\sim E-\langle V\rangle. \tag{15}$$

Then

$$\langle V\rangle=\int V(r)u^2(r)d\tau=\int_0^\infty V(r)P^2(r)dr,$$

and from the $P(1s)$ function given by Wilson and Lindsay,[3] we find

$$\langle V\rangle\sim0.0895 \text{ ry}$$

and for the integral in (4), 2.0522 ry. Therefore, the total value is

$$E'=-5.814 \text{ ry}.$$

This is to be compared to an experimental energy of

$$E_{\text{exp}}=-5.8067 \text{ ry}$$

(and the energy calculated with Hylleraas' 6-parameter wave function[8]: $E_{\text{Hyl}}=-5.8057$ ry).

Thus the correlation energy (-0.0895 ry), as calculated here, gives 108% of the actual error inherent in the Hartree approximation.

I wish to thank Professor Joseph Callaway, who suggested the problem, and him and Professor Eugene Wigner, for their guidance.

[6] J. C. Slater, Phys. Rev. **81**, 385 (1951).
[7] G. Pratt, Jr., Phys. Rev. **88**, 1217 (1952).
[8] Green, Mulder, and Milner, Phys. Rev. **91**, 35 (1953); P. Gombás, Acta Physica Hung. **4**, 187 (1954).

Isotope Shift in the First Spectrum of Atomic Lithium*

R. H. HUGHES†
Physics Department, University of Wisconsin, Madison, Wisconsin
(Received June 13, 1955)

The isotope shift in Li I has been studied in the hollow cathode spectrum from separated isotopes with the aid of a Fabry-Perot interferometer. Specific shifts have been determined for the $2s$, $2p$, $3s$, $3p$, $3d$, $4s$, $4d$, $5s$, and $5d$ levels. The inadequacy of the Hughes-Eckart approximation is pointed out.

I. RESUMÉ OF THEORY

THE isotope shift in the atomic spectra of light elements is considered to result exclusively, within the limits of measurements, from the finite mass of the nucleus. Hughes and Eckart[1] were the first to treat mass effects in atoms with more than one electron. Relativistic and nuclear structure effects being neglected, the expression for the kinetic energy of an atom with a nucleus of finite mass differs from that of an atom with the same charges and an infinite mass nucleus by the additional terms $(2M)^{-1}(\sum p_i^2 + \sum' p_i \cdot p_j)$, where M is the nuclear mass and p_i the momentum of the ith electron, and $\sum'$ means sum only over pairs $i \neq j$. The coordinates are measured relative to the nucleus.

The first sum results in the normal mass effect, which contracts the energy level spectrum of the atom with the finite mass nucleus by the ratio of the reduced mass to the electronic mass. The second sum is the specific mass effect, the expectation value of which must be calculated. Hughes and Eckart, using linear combinations of products of hydrogen-like wave functions as their approximation and neglecting interconfiguration mixing, calculated the specific mass effect for two- and three-electron spectra. In their approximation the specific shift occurs only for those configurations that have electrons differing by unity in l-value, e.g., among the lightest atoms, for p-terms only.

II. EXPERIMENTAL APPARATUS

The light sources used throughout were liquid air-cooled hollow cathode discharge tubes of the Mack-Arroe design.[2] Separated isotopes[3] with abundance 99.91 and 95.4 percent of Li^7 and Li^6, respectively, were used. Resolution was obtained by crossing a Fabry-Perot interferometer with a special Hilger constant-deviation spectrograph or a Bausch and Lomb medium quartz spectrograph.

III. TREATMENT OF DATA

In this paper, the whole shift of each level is assumed to consist of only two parts: the normal mass effect, which can be calculated with negligible uncertainty, and the specific mass effect. Thus, for each line, the observed total shift minus the calculated normal mass shift is considered to be the "observed specific mass shift", which must be the difference between the specific mass shifts of the two levels.

The following sign convention is used: For a line, the normal shift is positive; for a level, positive shift implies the relative depression of the heavier isotope, the matching point being the series limit, Li II $1s^2\ {}^1S_0$.

The magnitude of the specific shift of a level considered as a member of a series, is assumed to be a monotonically decreasing function of the principal quantum number, if the levels are arbitrarily matched at the series limit. (The behavior of the shifts in the line series $2p$–ns and $2p$–nd will be seen below to tend to justify this assumption.) The consequent asymptotic vanishing of the shift is used as a criterion to fix the otherwise undetermined zero point for the shift of the common level.

No account has been taken of the hyperfine structure splitting of the lines, which, being in every case less than the Doppler width, has not been observed and is believed not to have affected the measurements appreciably.

The presence of the weaker isotope (4.6 percent of Li^7 in Li^6, and 0.09 percent of Li^6 in Li^7) has been neglected; this we believe to be fairly safe, since none of our lines was overexposed.

Wave numbers are measured in kaysers, K (1K=1 cm^{-1}), and shifts in millikaysers, mK.

The errors in this work, indicated by the ± sign, are estimated limit (high-confidence) errors.

IV. RESULTS

The experimental results of the work, some of which have been reported upon,[4] are shown in Table I. In the transition $2s$–$2p$ at $\sigma = 14\,904$ K ($\lambda = 6708$ A) it was found, in disagreement with Jackson and Kuhn,[5] that within the limits of error of ±3 mK, there is no isotopic

* This work was done under the supervision of Professor J. E. Mack and was supported in part by the Office of Naval Research, the U. S. Office of Ordnance Research, and the C. E. Mendenhall Fellowship.
† Now at the Physics Department, University of Arkansas, Fayetteville, Arkansas.
[1] D. S. Hughes and C. Eckart, Phys. Rev. **36**, 694 (1930).
[2] O. H. Arroe and J. E. Mack, J. Opt. Soc. Am. **40**, 386 (1950).
[3] Obtained on loan from the Y-12 plant, Oak Ridge National Laboratory, on AEC allocation.
[4] R. H. Hughes, Phys. Rev. **91**, 457(A) (1953); **95**, 621(A) (1954).
[5] D. A. Jackson and H. Kuhn, Proc. Roy. Soc. (London) **A173**, 278 (1939).

TABLE I. Observed doublet separation and isotope shift of lithium lines (in mK).

Transition	σ(K)	Separation[a]	Observed shift[a]	Normal shift	Residual shift[a]	H[b]	Previously observed shift JK[c]	MMS[d]	Mc[e]
$2s$–$2p$	14 904	337 ±3	+350.$_9$± 3	+194.0	+156.$_9$± 3	+347	{+345$_1$ / +365$_1$}	(+356±2)$_1$	f
$2s$–$3p$	30 925	g	+482.$_7$±15	+402.5	+ 80.$_2$±15	+555			
$2p$–$3s$	12 302	336.$_3$±2	+ 30.$_6$± 2	+160.1	−129.$_5$± 2				
$2p$–$3d$	16 379	302.$_7$±3[g]	+ 92.$_6$± 3	+213.2	−121.$_2$± 3	h			
$2p$–$4s$	20 108	336 ±3	+140.$_2$± 3	+261.7	−121.$_4$± 3				
$2p$–$4d$	21 719	322.$_3$±4[g]	+163.$_4$± 4	+282.7	−118.$_9$± 4				
$2p$–$5s$	23 396	336 ±4	+186.$_1$± 4	+304.5	−118.$_4$± 4				
$2p$–$5d$	24 191	327.$_3$±4[g]	+195.$_6$± 4	+314.8	−119.$_6$± 4				

[a] The digit after the decimal point is carried along for computational purposes although its significance is slight.
[b] D. S. Hughes, Phys. Rev. **38**, 857 (1931).
[c] See reference 5.
[d] See reference 6.
[e] J. R. McNally, Jr., Am. J. Phys. **20**, 152 (1952).
[f] Identical with the doublet separation.
[g] Fine structure unresolved or only partly resolved.
[h] Undetected.

difference in the $2p$ doublet separation, i.e., the isotope shift has no dependence upon the total angular momentum, J. (In agreement with our finding, there appears to be no place in the theory for the fine structure to be so strongly mass-dependent.)

The specific shift data for the levels are gathered in Table II. The values in the line series $2p$–ns and $2p$–nd converge to −119±4 mK, which is therefore assigned to the $2p$-levels. With this assignment to start with, the specific shift values for the remaining levels are determined.

The observed specific s- and p-shifts are approximately $+1.4\times10^{-1}$ K$\times n_{eff}^{-2.8}$ and -6.4×10^{-1} K $\times n_{eff}^{-2.5}$, respectively, where n_{eff}=(Rydberg constant)[½] (term value)$^{-½}$. The exponents are equal, well within the limits of error. It is not evident *a priori* that such a simple relationship should occur.

The doublet separations in all our measured members of the sharp series, including the first member of the principal series, agree with the atomic-beam values of both Jackson and Kuhn[5] (337.2±0.5 mK) and Meissner, Mundie, and Stelson[6] (336.6±0.5 mK), within our larger (and differently specified—see Sec. III) uncertainty.

The $2p$–nd transitions confirm the values of Meissner, Mundie, and Stelson for the $3d$, $4d$, and $5d$ doublet separations. Although the ^2D structure was not resolved, the splittings can be calculated from the quantum-statistical weights as 38±4 mK, 16±5 mK, and 10±5 mK, respectively. Here Meissner's $2p$ doublet separation value of 336.6±0.5 mK has been used. The $3p$ ^2P structure was unresolved, a circumstance which accounts for the large uncertainty in the $3p$-shift.

TABLE II. Observed and calculated specific shift in Li I levels (in mK).

	ns ^2S		np ^2P			nd ^2D	
n	Obs	Calc HE[a]	Obs	HE[a]	Calc b	Obs	Calc HE[a]
2	+38±7	0	−119± 4	−80	−90		
3	+11±6	0	− 42±22		−24	+2±7	0
4	+ 3±7	0				0[c]	0
5	0[c]	0				0[c]	0

[a] See reference 1.
[b] Calculated with modern mass values of the atomic masses, and using Morse-Young-Haurwitz functions [Phys. Rev. **48**, 948 (1935)] (with Z_1, Z_3 extrapolated to 2.69, 1.00 for $3p$).
[c] Bases for the assignment of the $2p$-shift.

[6] Meissner, Mundie, and Stelson, Phys. Rev. **74**, 932 (1948); erratum, **75**, 891 (1948).

V. DISCUSSION

It would be difficult to find an explanation for the s-shifts outside of the specific mass effects. The field effects give a shift in the opposite direction, which is commonly expected to be very small in any case, for such light nuclei. (The nuclear volume effect would give the correct sign only if the Li^6 nucleus were the larger.) We are led to question the validity of the Hughes-Eckart approximation for the wave function. It is interesting to note the effect on the calculated specific shift for He I of improved wave functions. Fred, Tompkins, Brody, and Hamermesh[7] point out that, although their use gives finite values for s-shifts, they are at this stage hardly adequate to explain the experimental values.

Burke[8] has found a large departure from the Hughes-Eckart calculation in the resonance line $2s$–$2p$ of the lithium-like spectrum B III. The discrepancy is in the normal effect direction and can be at least partially explained by a positive shift in the $2s$ level.

The calculated specific shifts in the $2p$ and $3p$ levels in Li I are too small by factors of about 3/4 and 4/7, respectively, which would not bear out the expectation of the relative improvement of the Hughes-Eckart approximation with increasing n; but the large uncertainty in the $3p$-shift prevents a final conclusion now concerning this n-dependence.

The evidence is overwhelming against the adequacy of the Hughes-Eckart approximation. If the present

[7] Fred, Tompkins, Brody, and Hamermesh, Phys. Rev. **82**, 406 (1951).
[8] E. Burke, following paper [Phys. Rev. **100**, 1839 (1955)].

theory of the isotope shift in the light elements is complete, the agreement between calculation and experimental observation should improve as more adequate wave functions are used in the specific mass calculations. As pointed out by Mack *et al.*,[9] such studies of the specific mass effect should prove to be useful in checking the accuracy of the wave functions adopted for the light elements.

[9] Bernarda, Burke, Burnett, Hughes, and Mack, Rydberg Conference, Lunds University Årsskrift 50:21, 1954 (unpublished), p. 88.

Gailer[10] reports that three lines in the lithium spectrum, at 3303 A, 3650 A, and 3720 A, remain to be classified. We believe all three lines to be spurious. A probable explanation for at least two of these lines lies in the list of the persistent lines of the elements. The second member of the principal series of sodium lies at 3303 A and the ultimate line of Fe I (iron was present in Gailer's system), at 3720 A. The other line, at 3650 A, could possibly be from Hg I.

[10] K. Gailer, Ann. Physik. **126**, 583 (1949).

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Isotope Shift in the First Three Spectra of Boron*

EDWARD W. BURKE, JR.†
University of Wisconsin, Madison, Wisconsin

(Received June 13, 1955)

Isotope shifts have been observed in the first three spectra of boron, with separated isotopes. On the assumption that the total observed shift consists only of the (exactly calculable) normal mass effect and a specific mass effect, the results are as follows (a positive sign indicating a shift of B^{11} toward the violet relative to B^{10}):

	Total shift	$\frac{\Delta\sigma_{s,\text{obs}}}{\Delta\sigma_{s,\text{calc}}}$ (approx)
B I $1s^22s^22p\ ^2P_{\frac{1}{2},\frac{3}{2}}-1s^22s^23s\ ^2S_{\frac{1}{2}}$	$\begin{cases}-137\pm12\text{ mK}\\-139\pm\ 6\text{ mK}\end{cases}$	1.0 or 0.9,
B II $1s^22s2p\ ^1P_1-1s^22p^2\ ^1D_2$	$+867\pm26$ mK	1.3,
B III $1s^22s\ ^2S_{\frac{1}{2}}-1s^22p\ ^2P_{\frac{1}{2},\frac{3}{2}}$	$\begin{cases}+1170\pm120\text{ mK}\\+1210\pm\ 90\text{ mK}\end{cases}$	1.4,

where the calculated specific shift, $\Delta\sigma_{s,\text{calc}}$ is based on hydrogen-like wave functions of the Hughes-Eckart type.

I. INTRODUCTION

HUGHES and Eckart[1] inaugarated the theoretical study of the isotope shift in light polyelectronic atoms in 1930. Hughes[2] at the same time made an experimental determination for lithium, but the work was done without the benefit of separated isotopes and the experimental results were not sufficient for a close check with theory. The conclusion was that the agreement was reasonable. Hughes[3] has recently investigated the isotope shift in neutral lithium with separated isotopes and found that observations give specific shifts larger than the calculated values.

The current investigation of boron was initiated to consider another three-electron system, B III. The only lines available are those of the resonance doublet at 48 392/58 K (kaysers, $1\text{ K}=1\text{ cm}^{-1}=10^3$ mK), 2066/7 A, $1s^22s\ ^2S_{\frac{1}{2}}-1s^22p\ ^2P_{\frac{1}{2},\frac{3}{2}}$. The studies have been extended to repeat Mrozowski's[4] measurements, which he made with unseparated isotopes, on the B II line 28 966 K (3451 A), $1s^22s2p\ ^1P_1-1s^22p^2\ ^1D_2$, and the B I doublet at 40 040/24 K (2497/8 A), $1s^22s^22p\ ^2P_{\frac{1}{2},\frac{3}{2}}-1s^22s^23s\ ^2S_{\frac{1}{2}}$. The B II $1s^22p^2\ ^1D_2-1s^22s3p\ ^1P_1$ line was not excited in our sources. Calculations have been made for the B I lines by Opechowski and Devries[5] and by Vinti.[6] A calculation for the B II line has been made by Vinti.[6] Hughes' paper[3] contains a brief outline of the theory.

II. EXPERIMENTAL

For excitation of the B III spectrum a special high-excitation, sharp-line source was required. The resonance doublet has been previously observed only in the hot spark, but high voltages in a highly condensed discharge could not be used since that would have produced broadened lines. A compromise was effected with a condensed-hollow-cathode source. The tube itself was

* Supported by the Office of Naval Research and the Office of Ordnance Research.

† Now at King College, Bristol, Tennessee.

[1] D. S. Hughes and C. Eckart, Phys. Rev. **36**, 694 (1930); the last equation on page 697 should be labeled "(13)".

[2] D. S. Hughes, Phys. Rev. **38**, 857 (1931).

[3] R. H. Hughes, preceding paper [Phys. Rev. **99**, 1837 (1955)].

[4] S. Mrozowski, Z. Physik **112**, 223 (1939).

[5] W. Opechowski and D. A. Devries, Physica **6**, 913 (1939).

[6] J. P. Vinti, Phys. Rev. **58**, 879 (1940).

TABLE I. Isotope shifts in the spectra of boron (values in mK except where noted).

Line λ(A)	σ(K)	$\sigma_{obs}-\sigma_{tab}$(K)	Classification	$\Delta\sigma_{obs}$[b,c]	$\Delta\sigma_n$	$\Delta\sigma_{s,obs}$	$\Delta\sigma_{s,calc}$ Lower level	Upper level	Line	$\Delta\sigma_{s,obs}$[b] / $\Delta\sigma_{s,calc}$
2065.783	48 392.33	−0.3	B III $1s^22s\ ^2S_{\frac{1}{2}}-1s^22p\ ^2P^o_{\frac{1}{2}}$	+121₀ ± 9₀	+240	+97₀ ± 9₀	0	−671	+671	1.45 ±0.14
2067.238	48 358.28	−0.2	B III $^2S_{\frac{1}{2}}-\ ^2P^o_{\frac{1}{2}}$	+117₀ ±12₀	+240	+93₀ ±12₀	0	−671	+671	1.39 ±0.18
2496.8	40 040		B I $1s^22s^22p\ ^2P^o_{\frac{1}{2}}-1s^22s^23s\ ^2S_{\frac{1}{2}}$	−137 ±12	+199	−336 ± 12	−33₀[d]	0	−33₀[d]	1.02 ±0.04[d]
							−367[e]	0	−367[e]	0.92 ±0.03[e]
2497.7	40 024		B I $^2P^o_{\frac{1}{2}}-\ ^2S_{\frac{1}{2}}$	−139 ± 6	+198	−337 ± 6	−33₀[d]	0	−33₀[d]	1.02 ±0.02[d]
							−367[e]	0	−367[e]	0.92 ±0.02[e]
3451.298	28 966.39	+1.0	B II $1s^22s2p\ ^1P^o-1s^22p^2\ ^1D_2$	+867 ±26	+144	+723 ± 26	−305[e]	−872[e]	+567[e]	1.28 ±0.05

[a] The uncertainty shown does not include that of the calculation.
[b] Positive (negative) shift means, direction the same as (opposite to) that of the normal mass shift, i.e., B^{11} shifted to the violet (red) from B^{10}.
[c] The ± values are estimated limit (i.e., high-confidence) errors.
[d] See reference 5.
[e] See reference 6.

similar to one described by Tolansky[7] as a hot hollow cathode. A somewhat similar discharge has been used by Paschen[8] in 1923 and by Gartlein and Gibbs[9] in 1931, and more recently by Edlén and Glad,[10] and others. A condensed pulsating voltage was applied to the tube. Capacitances varying from $\frac{1}{4}$ to 1 μf were used. The condenser was charged to the desired voltage with a power supply and allowed to discharge through the tube with a mechanical interrupter consisting of a four-pronged wheel, rotating at about 75 rpm, making contacts with spring-brass strips. The discharge appeared steady and was concentrated mainly inside the cathode. Small sparks appeared within the cathode during operation. Optimum operation for the excitation of the B III resonance lines used a $\frac{1}{2}$-μf condenser at 1200 volts. Exposure times of ten minutes were sufficient with a 21-foot grating at f:42.

The vacuum system for the B III lines was small, simple and static. The single vessel contained two cathodes so prepared that the two separated isotopes of boron could be used alternately. When a forepump and an efficient charcoal trap had produced a black vacuum, helium was allowed to enter through a three-valve system until the brightest hollow cathode discharge was observed. The condensed spark could then be started.

Although the B II 28 966 K line and the B I 40 040/24 K lines could be excited in the above-described tube operating without the condensed spark, these lines with their small isotope shifts could not be studied in the hot hollow cathode because of Doppler broadening, so an Arroe-Mack tube[11] was employed. Two such tubes were mounted in parallel so that the separated isotopes could be photographed simply by switching, without otherwise changing the operating conditions. The vacuum, which again was static, was obtained as it was with the condensed hollow-cathode discharge system.

A Fabry-Perot interferometer in series with a quartz E-1 Littrow spectrograph was employed for the study of the shift in the B I 40 040/24 K doublet. All other studies were made with the 21-foot Rowland mount grating spectrograph.

The photographing of radiation above 47 000 K required special emulsions. Two were used in this study. The Eastman SWR plate proved satisfactory; however, the Eastman IIa-Ouv proved to be more convenient to handle and slightly more sensitive in that region. Eastman 103-0 plates were satisfactory for all the lines to the red of 47 000 K.

III. RESULTS

The results of this investigation may be most readily seen from Table I. Here σ_{tab} refers to the Moore-Sitterly[12] atomic energy level tables. The lines listed in the $(\sigma_{obs}-\sigma_{tab})$ column have been carefully remeasured, the B III lines at concentrations 96 percent B^{10} and 97 percent B^{11} respectively, in the fifth grating order against third-order iron lines with precautions in alignment to minimize the errors inherent in this method, and the B II lines in the third order against third-order iron lines. The normal mass shifts, $\Delta\sigma_n$, have been recalculated with modern values for the constants.[13,14] According to our sign convention, which is that of most investigations of the isotope shift in the spectra of light atoms, the normal shift in a line is positive, and for a level, positive shift means the relative depression of the heavier isotope, the matching point being a common series limit. The column labeled $\Delta\sigma_{s,obs}$ is simply the difference between the previous two columns, and carries the implied assumption that the total shift arises only from the normal and the specific mass shifts of the levels. In the $\Delta\sigma_{s,calc}$ columns, the value without a reference number is from our own calculations, made with the aid of Hughes and Eckart's[1] Eq. (13), with the values $Z_1=4.68$, $Z_2=3.19$ after Morse, Young, and Haurwitz.[15]

The isotope shift itself in the B III lines was measured

[7] S. Tolansky, *High Resolution Spectroscopy* (Pitman Publishing Corporation, New York, 1947), p. 50.
[8] F. Paschen, Ann. Physik **71**, 142 (1923).
[9] C. W. Gartlein and R. G. Gibbs, Phys. Rev. **38**, 1907 (1931).
[10] B. Edlén (private communication to J. E. Mack).
[11] O. H. Arroe and J. E. Mack, J. Opt. Soc. Am. **40**, 386 (1950).
[12] Charlotte E. Moore, Atomic energy levels, National Bureau of Standards Circular **467**, 1949, Vol. I.
[13] E. R. Cohen, Phys. Rev. **88**, 353 (1952).
[14] J. E. Drummond, Phys. Rev. **97**, 1004 (1955).
[15] Morse, Young, and Haurwitz, Phys. Rev. **48**, 948 (1935).

relative to neighboring aluminum lines on 15 plates; although the lines were much sharper than could be obtained from a hot spark, still they had an average half-intensity width of about 1300 mK, and were not easy to measure. The ratio of about 1.4 between observed and calculated values shows a need for restudy of the theory.

The B II $2s2p\ ^1P_1-2p^2\ ^1D_2$ line shift was completely resolved with natural boron, which has a B^{11}/B^{10} abundance ratio of about 4.3 or 4.4; however, better exposures for measurement were made by mixing the isotopes so that the concentrations were approximately equal. The value of $+867\pm26$ mK agrees with Mrozowski's[3] earlier value of $+877\pm4$. The error in our determination is a limit (high confidence) error obtained from 15 readings on each of two similarly exposed plates. In part of Mrozowski's paper the $\pm$ stands for twice the greatest departure of any measurement from the mean, but it is unlikely that the readings on this line could be so consistent. Here the ratio of almost 1.3 between the observed and the calculated specific shift again indicates rather poor agreement.

Fabry-Perot fringe systems were compared for the B I 40 040/24 K doublet, taken with separated isotopes at the concentrations used for the B III investigation. The values for the isotope shift of these lines are -137 ± 12 and -139 ± 6 mK, respectively. These values are believed to be better than Mrozowski's[4] earlier values of -175 ± 12 and -168 ± 10 mK taken from microphotometer traces of asymmetries in Fabry-Perot fringes from natural-abundance samples, with Doppler half-intensity breadths about equal to the line separation. Excellent agreement exists between the observed shift and the specific shift of -330 mK calculated for these lines by Opechowski and Devries[6] with Hartree radial functions. Vinti's[6] recalculation, using Morse, Young, and Haurwitz[15] wave functions yielded -366 mK, appreciably higher than the value reported here and in close agreement with Mrozowski's values, which with the use of our value for the normal shift, come to -373 ± 12 mK.

IV. CONCLUSIONS

In B III the great practical advantage of having an extraordinarily large specific isotope shift is largely offset by the considerable line width occasioned by the necessity of exciting the third spectrum. It is unusual to find a specific mass shift so large compared with the normal mass shift as occurs in our B III and B II lines. The considerable disagreement in a ratio of about 1.4 between the observed and the calculated values in the B III resonance lines indicates that the wave functions without interconfiguration interaction, used thus far in the calculations, are inadequate. Hughes'[9] values for the same transition ($1s^22s-1s^22p$) in Li I show almost as high a ratio (1.32 ± 0.04) between the observed and the calculated values; moreover, he was able to show that there must be a positive isotope shift in the ground level. It appears reasonable to assume a corresponding shift in the B III ground level. This can be understood if an admixture with $1s2p^2$ is assumed in the ground level. The more complicated configurations in B II and B I show successively smaller disagreement.

V. ACKNOWLEDGMENTS

The author would like to express his appreciation to Professor Julian Ellis Mack for the encouragement, aid, and advice he provided while supervising this work. He would also like to acknowledge the advice and aid freely given by Dr. Hack Arroe and Mr. J. A. Vreeland.

Two-Quantum Transitions in the Microwave Zeeman Spectrum of Atomic Oxygen*

V. W. HUGHES AND J. S. GEIGER†‡
Yale University, New Haven, Connecticut
(Received May 20, 1955)

Two-quantum transitions have been observed in the microwave Zeeman spectrum of atomic oxygen in its ground $^3P_2$ state by the method of magnetic resonance absorption spectroscopy. The three lines observed originally by Rawson and Beringer are identified as arising from two-quantum transitions between Zeeman levels with the selection rule $\Delta M = \pm 2$. Each of the three lines is observed at the mean frequency of the two corresponding $\Delta M = \pm 1$ transitions to within the experimental accuracy of a few parts per million. The line width of the two-quantum transitions is approximately one-half that of the normal transitions, and the line intensity varies more rapidly with rf power, both in agreement with the theory. The principal features of these transitions are explained by second-order time-dependent perturbation theory. The two-quantum transitions reported here are essentially similar to the double-quantum transition reported by Hughes and Grabner in the electric quadrupole spectrum of $Rb^{85}F$ and to the multiple-quantum transitions seen by Kusch in the Zeeman spectra of K and O_2.

1. INTRODUCTION

A MULTIPLE quantum transition between two atomic or molecular energy levels is a transition in which the energy is supplied by two or more quanta. The transition is forbidden in first order perturbation theory and occurs through one or more intermediate states. In recent years, several examples of multiple-quantum transitions have been reported in atomic and molecular beam experiments. A double-quantum (or half-frequency) transition was first observed in the electric quadrupole spectrum of $Rb^{85}F$ by Hughes and Grabner using the molecular beam electric resonance method.[1] The transition was interpreted by second order perturbation theory, according to which two equienergetic quanta, each of one half the Bohr frequency for the transition, supply the energy for the transition.[2] Also it was pointed out that the line width for a double-quantum transition should be one-half that of a single-quantum transition in agreement with the experimental observation. Subsequently, it was shown experimentally and theoretically that the transition occurs also if two different frequencies are applied, provided that the sum of the two frequencies equals the Bohr frequency for the transition.[3] Recently Kusch observed double- and triple-quantum transitions in the Zeeman spectrum of K and O_2[4]; Braunstein and Trischka observed double-quantum transitions in the Stark spectrum of LiF molecules[5]; Hamilton *et al.* observed multiple-quantum transitions in the hyperfine spectrum of Au^{198}.[6] The theory of transitions involving the emission or absorption of two quanta (processes closely related to the Raman effect) was first discussed by M. Goeppert Mayer on the basis of the Dirac radiation theory.[7] Recently Salwen has discussed the line shapes of multiple-quantum transitions, including the effect of the velocity distribution in an atomic beam experiment.[8]

It is the purpose of this paper to point out an example of a double-quantum transition in the Zeeman spectrum of atomic oxygen observed in a microwave magnetic resonance absorption spectroscopy experiment. The double-quantum transitions being reported here were observed originally by Rawson and Beringer but remained unidentified.[9] Further, a slight variation of the second order perturbation theory originally proposed to explain double-quantum transitions in electric quadrupole spectra[2] is applied to double-quantum transitions in Zeeman spectra of the type observed by Kusch and reported here.

A preliminary report of this work has already appeared.[10]

2. EXPERIMENTAL DATA

The ground state of atomic oxygen is a $(2p)^4$ configuration with three fine-structure levels as shown in Fig. 1. In the presence of a magnetic field the $^3P_2$ level is split into five magnetic sublevels designated by the magnetic quantum number M with values from -2 to $+2$. Because of incipient Paschen-Back effect, i.e., the mixing in of $^3P_1$ state by the magnetic field, adjacent magnetic sublevels of the $^3P_2$ state are not equally spaced, and hence the magnetic resonance spectrum consists of four lines.

The method of observation was the microwave magnetic resonance absorption method developed by Beringer and his students,[11] and the apparatus was that

* This research has been supported in part by the Office of Naval Research.
† To be submitted by J. S. Geiger in partial fulfillment of the Ph.D. thesis requirement at Yale University.
‡ Loomis-Sheffield Fellow, 1954–55.
[1] V. W. Hughes and L. Grabner, Phys. Rev. **79**, 314 (1950).
[2] V. W. Hughes and L. Grabner, Phys. Rev. **79**, 828 (1950).
[3] L. Grabner and V. W. Hughes, Phys. Rev. **82**, 561 (1951).
[4] P. Kusch, Phys. Rev. **93**, 1022 (1954).
[5] R. Braunstein and J. W. Trischka, Phys. Rev. **98**, 1092 (1955); private communication from J. W. Trischka on Li^6F^{19}.
[6] Private communication from D. R. Hamilton.
[7] M. Goeppert Mayer, Ann. Physik **9**, 273 (1931). See also G. Breit, Revs. Modern Phys. **4**, 504 (1932).
[8] H. Salwen, Phys. Rev. **99**, 1274 (1955).
[9] E. B. Rawson and R. Beringer, Phys. Rev. **88**, 677 (1952).
[10] V. W. Hughes and J. S. Geiger, Bull. Am. Phys. Soc. **30**, No. 3, 66 (1955).
[11] R. Beringer and J. G. Castle, Phys. Rev. **78**, 581 (1950).

used by Beringer and Heald.[12] A spectrum of the type originally studied by Rawson and Beringer[9] is shown in Fig. 2. Essentially the derivative of the absorption is plotted as a function of the magnetic field for a fixed microwave frequency. The magnetic field was measured with the so-called regulator proton resonance probe,[12] which was placed adjacent to the microwave cavity. The inclined line indicates the zero drift of the galvanometer output and line centers are at the cross-over points on this line. The four spectral lines *a*, *b*, *c*, and *d* are the expected Zeeman transitions corresponding to the transitions indicated in Fig. 1. These lines have been shown to occur to within the experimental accuracy[9] of a few ppm (parts per million) at the fields predicted by elaborate Zeeman calculations of Abragam, Van Vleck, and Kambe[13] which take into account relativistic and quantum electrodynamic effects. The extra lines *e*, *f*, and *g* remained unidentified by Rawson and Beringer. The suggestion was made in their work that the extra lines might be due to argon impurity in the tank oxygen, since the metastable $^3P_2$ state of argon also has a *g*-value of about 1.5. However, lines of the type shown in Fig. 2 have been taken with oxygen which was 99.6% pure (Matheson Company extra dry oxygen), and any appreciable line intensity from an impurity could not be expected.

Inspection indicates first that to within the experimental accuracy of a few ppm the magnetic field at which line *e* occurs is $\frac{1}{2}$ of the sum of the fields at which lines *a* and *b* occur. Similarly, line *f* occurs at the mean field of lines *b* and *c*, and line *g* occurs at the mean field of lines *c* and *d*. Since the principal part of the Zeeman energy depends linearly on the magnetic field *H*, it can be concluded to sufficient accuracy that for a fixed magnetic field the frequency of line *e* would be $\frac{1}{2}$ of the sum of the frequencies of lines *a* and *b*, and similar statements apply for lines *f* and *g*. The second observation is that the widths of the unidentified lines *e*, *f*, and *g* are considerably less than—indeed approximately $\frac{1}{2}$ that of—the normal lines *a*, *b*, *c*, and *d*.

Figure 3 presents information on the relative line intensities of the extra and normal lines as a function of microwave power measured at the bolometer. It is noticed that the intensity of the extra lines decreases considerably more rapidly with decrease of microwave power than does the intensity of the normal lines for the range of powers used.

Finally, a search indicated no additional unexpected lines in the oxygen spectrum. The two lines associated with the $^3P_1$ Zeeman splittings were seen, of course, as has been reported in the work of Rawson and Beringer.[9]

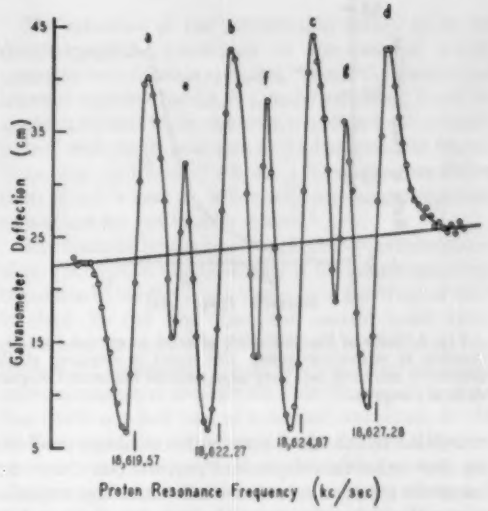


FIG. 2. Paramagnetic resonance absorption spectrum of ground state $^3P_2$ oxygen. Essentially the derivative of the absorption is plotted *vs* the magnetic field for a fixed microwave frequency of 9188.528 Mc/sec. The mean fields of single-quantum transitions and the fields of the extra lines are as follows:

$\frac{1}{2}(a+b) = 18\,620.92$ kc/sec; $e = 18\,620.93$ kc/sec,
$\frac{1}{2}(b+c) = 18\,623.57$ kc/sec; $f = 18\,623.58$ kc/sec,
$\frac{1}{2}(c+d) = 18\,626.08$ kc/sec; $g = 18\,626.08$ kc/sec.

The magnetic field was measured with the regulator proton resonance probe, which is placed adjacent to the microwave cavity. An additive correction would have to be applied to obtain the magnetic field in the cavity. This correction would not affect significantly the foregoing result that the extra lines occur at the mean field of corresponding single-quantum transitions. This curve was taken with ordinary tank oxygen.

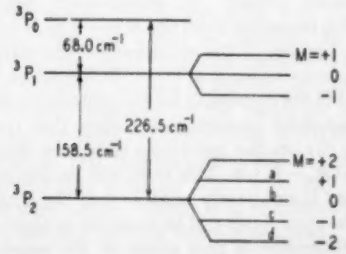


FIG. 1. Energy levels of atomic oxygen in a magnetic field.

3. THEORY OF TWO-QUANTUM TRANSITIONS

Consider an atomic state with three equally spaced Zeeman energy levels as, for example, the $^3S_1$ state of helium. The Zeeman energies are given by $W = \mu_0 g_J H M$, in which the symbols have their usual significance and $M = +1$, 0, or -1. It should be emphasized that the energy spacings between adjacent Zeeman energy levels are equal. It is well known that if such an atom is initially in the state $M = -1$ and a radio-frequency magnetic field with a component perpendicular to the static field and a frequency equal to $\mu_0 g_J H/h$ is applied, then after a time *t* the atom will have a certain probability of being in the states $M = 0$ or $M = +1$.[14] The

[12] R. Beringer and M. A. Heald, Phys. Rev. **95**, 1474 (1954).
[13] A. Abragam and J. H. Van Vleck, Phys. Rev. **92**, 1448 (1953); K. Kambe and J. H. Van Vleck, Phys. Rev. **96**, 66 (1954).
[14] V. W. Hughes *et al.*, Phys. Rev. **91**, 828 (1953).

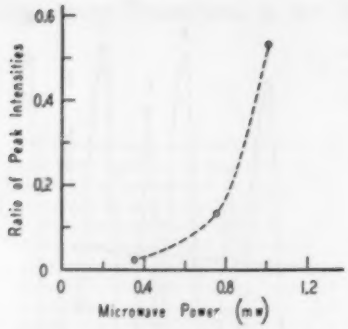


FIG. 3. Ratio of line intensities of extra to normal lines as a function of microwave power. The power is measured at the bolometer, and these data were taken with the Matheson Company extra dry oxygen.

transition to the $M=0$ state is the usual one predicted by first order time-dependent perturbation theory because the selection rule allows $\Delta M=\pm 1$. The transition to the $M=+1$ state occurs by virtue of the intermediate state $M=0$, and, since the energy levels are equally space, the usual interpretation would be that a resonance transition occurs from the $M=-1$ to $M=0$ state with the absorption of one quantum, and then a second resonance transition occurs from the $M=0$ to the $M=+1$ state with the absorption of a second quantum. The transition probability is given by the well known Majorana formula.[15]

The Zeeman spectrum of atomic oxygen presents a different situation because adjacent energy levels are not equally spaced, as was pointed out in the previous section. Thus, for example, if we consider an atom in the $M=-2$ magnetic substate of the $^3P_2$ level and apply a radio-frequency magnetic field whose frequency is the Bohr frequency for the transition from $M=-2$ to $M=-1$ the atom will, of course, have a substantial probability of making the transition to the $M=-1$ state, but the frequency will be far off resonance for a transition from the $M=-1$ to the $M=0$ state, so there will be negligible probability for the atom to reach the $M=0$ state.

It will now be pointed out that by a consideration of second-order time-dependent perturbation theory a substantial probability for the transition from the $M=-2$ to the $M=0$ state of $^3P_2$ oxygen is expected at a frequency one half the Bohr frequency for the $M=-2$ to $M=0$ transition. The Hamiltonian is:

$$\mathcal{H}=\mathcal{H}_0+\mathcal{H}',$$
$$\mathcal{H}'=\mu_0 g_J \mathbf{J}\cdot\mathbf{H}_{rf}=\mathcal{H}'(0)e^{-i\omega t};\quad \mathbf{H}_{rf}=\mathbf{H}_1 e^{-i\omega t},$$

in which $\mathcal{H}_0$ is the time-independent part of the Hamiltonian which leads to Zeeman levels as shown in Fig. 1. $\mathcal{H}'$ is the interaction with the rf magnetic field, and in this expression μ_0 is the Bohr magneton, g_J is the atomic g-value, $\mathbf{J}$ is the total atomic angular momentum operator, and $\mathbf{H}_{rf}$ is the applied radio-frequency magnetic field.[16]

First-order time-dependent perturbation theory gives the probability amplitude that an atom initially in the state n shall be in the state m after it has been subjected to the radio-frequency field for a time t [see Eq. (1)].

$$a_m^{(1)}(t)=(-i/\hbar)(m|\mathcal{H}'(0)|n)\times\frac{\{\exp[-i\omega t+i(E_m-E_n)t/\hbar]-1\}}{[-i\omega+i(E_m-E_n)/\hbar]}\simeq-(it/\hbar)(m|\mathcal{H}'(0)|n),\quad(1)$$

The usual resonance denominator appears for ω near the Bohr frequency. Provided t is sufficiently small so that the initial state amplitude can still be considered approximately 1, the expression is approximately that given on the right. Similarly, second-order time-dependent perturbation theory gives the probability amplitude:

$$a_m^{(2)}(t)=-\frac{1}{\hbar^2}\frac{(m|\mathcal{H}'(0)|l)(l|\mathcal{H}'(0)|n)}{[-i\omega+i(E_l-E_n)/\hbar]}\times\left\{\frac{\exp[-i2\omega t+i(E_m-E_n)t/\hbar]-1}{[-i2\omega+i(E_m-E_n)/\hbar]}-\frac{\exp[-i\omega t+i(E_m-E_l)t/\hbar]-1}{[-i\omega+i(E_m-E_l)/\hbar]}\right\}\simeq\frac{it}{\hbar^2}\frac{(m|\mathcal{H}'(0)|l)(l|\mathcal{H}'(0)|n)}{[-\omega+(E_l-E_n)/\hbar]},\quad(2)$$

where l designates an intermediate state.

Consider the case realized for the $^3P_2$ magnetic substates of oxygen because of the unequal spacing of adjacent Zeeman levels, in which the applied frequency ω can be far off resonance from the Bohr frequencies for either the $M=-2$ to $M=-1$ or the $M=-1$ to $M=0$ transitions but approximately equal to one-half the Bohr frequency for the transition $M=-2$ to $M=0$. If the atom is initially in the $M=-2$ state, there is negligible transition probability to the $M=-1$ state which is the only allowed transition in first-order perturbation theory. In second-order perturbation theory the transition from the $M=-2$ to $M=0$ state is allowed, because of the occurrence of the intermediate state l corresponding to $M=-1$. Furthermore, it will be observed in the expression for the probability amplitude in second-order perturbation theory that one of the terms in the bracket involves a resonance denominator for $2\omega=(E_m-E_n)/\hbar$. This term will be dominant, and provided t is sufficiently small so that the amplitude of the initial state n can still be regarded as approximately 1, the expression is that given on the second line of Eq. (2). The occurrence of 2ω instead of ω in the

[15] E. Majorana, Nuovo cimento 9, 43 (1932); F. Bloch and I. I. Rabi, Revs. Modern Phys. 17, 237 (1945).

[16] As usual it is only necessary to consider a single rotating component of the radiofrequency field. F. Bloch and A. Siegert, Phys. Rev. 57, 522 (1940).

resonance denominator implies that two equienergetic quanta are involved in the transition. It also implies that the line width for these two-quantum transitions will be $\frac{1}{2}$ that of the normal transitions. It was pointed out above that this was one of the characteristics of the unidentified lines.

The principal surprise about these two-quantum transitions is their substantial intensity relative to the normal transitions. This fact can be understood by reference to Eqs. (1) and (2). It will be noticed that the probability amplitude for a two quantum transition at its resonance frequency differs from the probability amplitude for a single-quantum transition at its different resonance frequency by a factor which is the ratio of a matrix element of the interaction with the rf field to the energy difference between the half-frequency for the transition from $M=-2$ to $M=0$ and the Bohr frequency for the transition from $M=-1$ to $M=0$. In the experimental situation the rf magnetic field was of the order of 0.1 gauss and the energy difference between the half-frequency and the $M=-1$ to $M=0$ frequency corresponds to about 0.28 gauss. Thus the ratio is of the order of 1 and hence the intensity of a two-quantum transition is comparable to that of a single-quantum transition. It will be noticed, however, that the intensity of a two-quantum transition should decrease more rapidly with decrease of rf field than the intensity of a single-quantum transition. This prediction is in agreement with the experimental results shown in Fig. 3. It should be emphasized that these remarks on line intensities are only intended in a qualitative sense.[17] The actual experimental conditions are such that the transition probabilities are high so that perturbation theory is not a very good approximation. A better knowledge of the experimental factors influencing line intensity such as relaxation phenomena and saturation effects, and a careful comparison with Salwen's theory[8] would be required for a more quantitative understanding of the line intensities.

Lines *e*, *f*, and *g* are, of course, being identified with the double-quantum transitions from $M=0\rightarrow M=+2$, $M=-1\rightarrow M=+1$, and $M=-2\rightarrow M=0$, respectively. The question arises about the double-quantum transition predicted in connection with the $^3P_1$ Zeeman spectrum. This line would be expected to be much weaker than the double-quantum transitions associated with the $^3P_2$ levels, because of the greater difference between the half-frequency for the transition $M=-1$ to $M+1$ and the Bohr frequency from $M=0$ to $M=+1$. Furthermore, this half-frequency line is predicted to occur at the field value for which a normal line in the $^3P_2$ spectrum occurs. Hence it is understandable that this double-quantum line was not observed.

[17] E. B. Rawson reports in his thesis [Yale, 1952 (unpublished)] that the relative intensity of the unidentified lines to the normal lines depends on pressure. Indeed a decrease in pressure caused this relative intensity to change from a value of about $\frac{1}{2}$ to a value of about 3. Possibly relaxation phenomena affect the double-quantum lines and the single-quantum lines differently, but no careful study has been made of this question.

The extension of the perturbation theory given for double-quantum transitions to the case of triple-quantum transitions is apparent.[18] Since the separations between adjacent lines *a*, *b*, *c*, and *d* are equal, it will be appreciated that triple-quantum transitions will coincide in field with single-quantum or double-quantum transitions [e.g., $(a+b+c)/3=b$, etc.]. This situation differs from Kusch's case in which adjacent single-quantum transitions are not equally spaced.[4]

The essential similarity of the double-quantum transitions discussed in this paper and of the double-quantum transition in the electrical quadrupole spectrum of RbF is clear. In the RbF case also second order time-dependent perturbation theory explains the basic features of the transition—the necessity of an intermediate state, resonance at one-half the Bohr frequency, and the line width one-half that of a normal transition. In the RbF example, the normal electric quadrupole transition can occur only when a static electric field is present, and the two-quantum transition can occur in the absence of the static electric field. The factor by which the probability amplitude for a two-quantum transition differs from that of a single-quantum transition is the ratio of the matrix element of the interaction with the radio-frequency electric field to the matrix element of the interaction with the static electric field. The ratio is of the order of 1 when the radio-frequency electric field is nearly equal to the static electric field.

The microwave magnetic resonance absorption method has certain advantages compared with conventional atomic beam experiments for a further study of the line intensities of multiple-quantum transitions as given in Salwen's theory: first, a more definite knowledge of the microwave field in a cavity than in an "rf-hairpin" of the type used in atomic beams, and, secondly, perhaps a better knowledge of the velocity distribution of the atoms, since no loss of low velocity atoms is involved as in the atomic beam case.

Finally, it might be pointed out that in conventional microwave spectroscopy involving transitions between different rotational states of a molecule, the conditions for the occurrence of double-quantum lines are not met because the separations of adjacent rotational levels differ by large factors. It seems possible that in the field of nuclear paramagnetic resonance where the level spacings are equal, one might find appropriate perturbations in some cases so that the conditions for a double- or multiple-quantum transition of the type discussed in this paper might be realized.

It is a pleasure to thank Professor R. Beringer for several helpful discussions about the apparatus.

[18] The principal term characteristic of a triple-quantum transition is found by third-order time-dependent perturbation theory to be:

$$a_m^{(3)}(t)=+\left(\frac{i}{\hbar^3}\right)\frac{(m|\mathcal{H}'(0)|k)(k|\mathcal{H}'(0)|l)(l|\mathcal{H}'(0)|n)}{[-i3\omega+i(E_l-E_n)/\hbar][-i2\omega+i(E_k-E_n)/\hbar]}\times\frac{\exp[-i3\omega t+i(E_m-E_n)t/\hbar]-1}{[-i3\omega+i(E_m-E_n)/\hbar]},$$

where *l* and *k* designate intermediate states.

Oscillator Strengths for Transitions between $2p^3 3p$ and $2p^3 3s$ in Oxygen I, Including Configuration Interaction*

R. F. KINGSBURY†
University of Pennsylvania, Philadelphia, Pennsylvania
(Received May 27, 1955)

Oscillator strengths are calculated for transitions between the various terms of $2p^3 3p$ and $2p^3 3s$ of oxygen I. Term energies for the two configurations are calculated with the inclusion of the $L(L+1)$ correction of Trees and with configuration interaction. The configurations included with $2p^3 3s$ are $2p^3 4s$ and $2s2p^5$ while $2p^3 4p$ is included with $2p^3 3p$. It is shown that the $L(L+1)$ correction should be assigned to the parent ion. Configuration interaction is included in the calculation of σ^2 as well as in the energies. The radial wave functions used in the calculation of σ^2 are Hartree wave functions without exchange.

INTRODUCTION

IT has been pointed out by Condon and Shortley[1] that one should expect configuration interaction to play a significant role in the determination of the energy levels for oxygen I since the terms of the series based on different parent terms of the ground state of oxygen II lie close together. For this reason, if one is to calculate oscillator strengths for transitions between excited states of O I, the wave functions for the various states should be linear combinations of the wave functions for the separate configurations. The proper linear combination will be that obtained by the unitary transformation which will diagonalize the energy matrix. In this work, oscillator strengths for transitions between the odd-parity levels predominantly of the $2p^3 3s$ configuration and the even levels, for which $2p^3 3p$ is the most important configuration, have been calculated.

OSCILLATOR STRENGTHS

The oscillator strength, f, for a transition between the levels γSLJ and $\gamma'S'L'J'$, defined in TAS 9[1], may be written in the form

$$f(\gamma SLJ,\gamma'S'L'J') = \frac{\nu(\gamma SLJ,\gamma'S'L'J')S(\gamma SLJ,\gamma'S'L'J')}{3R(2J+1)a_0^2e^2}, \quad (1)$$

where $\nu(\gamma SLJ,\gamma'S'L'J')$ is the wave number difference between the two levels, R is the Rydberg constant expressed in wave numbers, a_0 is the radius of the first Bohr orbit, $2J+1$ is a weighting factor for the initial level, and $S(\gamma SLJ,\gamma'S'L'J')$ is a parameter called the line strength for the transition between the two levels. The line strength for a line is the sum of the strengths of the components of the line:

$$S(\gamma SLJ,\gamma'S'L'J') = \sum_M \sum_{M'} s(\gamma SLJM,\gamma'S'L'J'M'), \quad M'=M,\ M\pm 1. \quad (2)$$

For electric dipole radiation, the line strength of the component of a line is the square of the matrix element of the electric dipole moment connecting the two states involved in the transition.

$$s(\gamma SLJM,\gamma'S'L'J'M') = \left|\int \Psi(\gamma SLJM)\mathbf{P}\Psi(\gamma'S'L'J'M')d\tau\right|^2. \quad (3)$$

Line strengths may also be written in the form[2]

$$S = S_1(L)S_2(M)\sigma^2, \quad (4)$$

where $S_1(L)$ is a factor which depends on the particular line of a multiplet and $S_2(M)$ depends on the particular multiplet of the transition array under consideration. The quantity σ is the integral $[a_0e/(4l^2-1)^{\frac{1}{2}}] \times \int R_{nl}(\rho)\rho R_{n,l-1}(\rho)d\rho$, where $R(\rho)/\rho$ is the radial wave function of the jumping electron in atomic units. $S_1(L)$ may be obtained from tables given either by White and Eliason[3] or by Russell,[4] while tables for $S_2(M)$ are given by Goldberg.[5,6]

It has been pointed out by Chandrasekhar[7] and others[8,9] that, where $R(\rho)$ is a solution of the Schrödinger equation, the dipole moment, dipole velocity, and dipole acceleration forms for the oscillator strength are equivalent. When, however, the wave functions are chosen on the basis of a variational calculation for the energy, these wave functions are most exact at moderate radial distances. This leads to a preference for the dipole velocity form. Since the radial wave functions used in these calculations were obtained by the Hartree method,[10] the integrals which were actually calculated were the dipole velocity integrals.

* This paper is part of a thesis presented to the University of Pennsylvania in partial fulfillment of the requirements for a Ph.D. degree.

† Now at Trinity College, Hartford, Connecticut.

[1] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge University Press, Cambridge, 1935) (This shall hereafter be referred to as TAS) 2[8].

[2] D. R. Bates and A. Damgaard, Trans. Roy. Soc. (London) **A242**, 101 (1949).

[3] H. E. White and A. Y. Eliason, Phys. Rev. **44**, 753 (1933).

[4] H. N. Russell, Astrophys. J. **83**, 129 (1936).

[5] L. Goldberg, Astrophys. J. **82**, 1 (1935).

[6] L. Goldberg, Astrophys. J. **84**, 11 (1936).

[7] S. Chandrasekhar, Astrophys. J. **102**, 223 (1945).

[8] L. C. Green and N. E. Weber, Astrophys. J. 111, 582 (1950).

[9] M. N. Lewis, National Bureau of Standards Report No. 2457, 1953 (unpublished).

[10] D. R. Hartree, Proc. Cambridge Phil. Soc. **24**, 89 (1928).

$$\int_0^\infty R_{n,l}(\rho)\rho R_{n,l-1}(\rho)d\rho = \frac{2}{\Delta\epsilon}\int_0^\infty R_{n,l-1}(\rho)\left[\frac{l}{\rho}R_{n,l}(\rho)+\frac{dR_{n,l}(\rho)}{d\rho}\right]d\rho. \quad (5)$$

When the initial and final states can be assigned uniquely to single configurations, σ^2 will be common to all transitions between the configurations. When, however, this unique assignment is impossible, σ^2 will depend on the pair of terms which are involved in the transition. The wave functions for the various states will be linear combinations of the individual configuration wave functions.

$$\Psi(\gamma SLJM)=a_1\Psi_1(\gamma_1 SLJM)+a_2\Psi_2(\gamma_2 SLJM)+\cdots. \quad (6)$$

The formula for the line strength will then be of the form

$$s(\gamma SLJM,\gamma' S'L'J'M') = \left|\int \sum_i \sum_{i'} a_i a_{i'} \psi_i(\gamma_i SLJM) \times \mathbf{P}\psi_{i'}(\gamma' SL'J'M')d\tau\right|^2. \quad (7)$$

The dependence of each of the integrals on L, J, M, L', J', and M' will be the same as it would have been if the states were uniquely assigned to configurations. $S_1(L)$ and $S_2(M)$ will, therefore, be the same for each integral but the σ^2 will be different and will be of the form

$$\sigma^2=(\sum_i \sum_{i'} a_i a_{i'} I_{ii'})^2. \quad (8)$$

In order to determine the proper linear combination of configuration wave functions, the matrices for electrostatic interaction were calculated, including the off-diagonal elements connecting the different configurations. In the case of the p^3p matrix, there are, in addition, off-diagonal elements connecting terms in the same configuration, having the same term designation, but built on different terms in the parent ion. These matrices were diagonalized, and the radial integral parameters were adjusted to give the best fit between theoretical and experimental[11] term values. When this had been done, the unitary transformation matrix which would perform the diagonalization was determined. This unitary matrix determines the proper linear combination of configuration wave functions.

In the next two sections, the procedures used to diagonalize the two energy matrices will be discussed in some detail.

ODD-PARITY CONFIGURATIONS

The configurations that are most likely to interact with the $2p^33s$ configuration are those with terms which lie nearest to the terms of $2p^33s$. The configurations included here are $2p^34s$ and $2s2p^5$. The diagonal matrix elements for a p^3s configuration are given in TAS 5[7] in terms of three parameters, E_0, F_2, and G_1. They were checked and it was found that they were all correct, except that for ^5S the coefficient of G_1 should be $+1$ instead of -1. This correction agrees with the relations as given by Johnson[12] and by Edlén.[13] The $2s2p^5$ configuration has but two terms and two parameters, E_0 and G_1. The off-diagonal elements of the matrix contain four additional parameters. The matrix elements are given in Table I. There are, therefore, in all twelve different parameters which can be adjusted in order to fit the experimental term values. F_2 involves only $2p$ electrons, which are assumed to remain unchanged, and this parameter was considered to be the same for both configurations $2p^33s$ and $2p^34s$. Of the fourteen possible term values eleven are known experimentally.[11] The other three were determined by extrapolation of the O I isoelectronic sequence.

TABLE I. Matrix elements of odd-parity configurations.

Term	Diagonal elements		Off-diagonal elements
$(^4S)ns\ ^5S$	$E_0(ns)-15F_2-3G_1(ns)$		R_0-3R_1
$(^4S)ns\ ^3S$	$E_0(ns)-15F_2+G_1(ns)$		R_0+R_1
$(^2D)ns\ ^5D$	$E_0(ns)-6F_2-2G_1(ns)$		R_0-2R_1
$(^2D)ns\ ^1D$	$E_0(ns)-6F_2$		R_0
$(^2P)ns\ ^3P$	$E_0(ns)$	$-2G_1(ns)$	R_0-2R_1
$(^2P)ns\ ^1P$	$E_0(ns)$		R_0
$2s2p^5\ ^3P$	$E_0(sp^5)$	$-2G_1(sp^5)$	$\sqrt{2}R_1'$; $\sqrt{2}R_1''$
$2s2p^5\ ^1P$	$E_0(sp^5)$		$\sqrt{2}R_1'$; $\sqrt{2}R_1''$

$E_0(ns)=3F_0(2p,2p)+3F_0(2p,ns)+$integrals over closed shells,
$E_0(sp^5)=10F_0(2p,2p)+5F_2(2p,2s)-20F_2(2p,2p)-G_1(2p,2s)$ +integrals over closed shells,
$R_0=3R_0(2p3s,2p4s)+$integrals over closed shells,

$F_2=F_2(2p,2p)$, $R_1=R_1(2p3s,4s2p)$,
$G_1(ns)=G_1(2p,ns)$, $R_1'=R_1(2s3s,2p2p)$,
$G_1(sp^5)=G_1(2p,2s)$, $R_1''=R_1(2s4s,2p2p)$.

The parameters were evaluated in a step-by-step process. First, a least-squares calculation was made using the diagonal sums of the submatrices. This calculation gave values for $E_0(sp^5)$, $G_1(sp^5)$ and F_2 and the sums $E_0(3s)+E_0(4s)$ and $G_1(3s)+G_1(4s)$. Next, a least-squares calculation was performed using the four two by two submatrices to determine $E_0(3s)$, $E_0(4s)$, $G_1(3s)$, $G_1(4s)$, R_0, and R_1. Lastly, a least-squares calculation was made with the two three by three submatrices to determine R_1' and R_1''. The results of these calculations are given in Table II, column *b*. The fit for the ^3P and ^1P terms is not at all satisfactory. Several other procedures were tried, but none gave satisfactory results although it was possible to reduce

[11] C. E. Moore, National Bureau of Standards Circular No. 467, 1949 (This is in book form under the title *Atomic Energy Levels*.)

[12] M. H. Johnson, Phys. Rev. **39**, 197 (1931).

[13] B. Edlén, Kgl. Svenska Vetenskapsakad. Handl. (3) **20**, No. 10 (1943).

TABLE II. Term values of $2p^33s$, $2p^34s$, and $2s2p^5$, in kaysers. (1 K = 1 cm^{-1}.)

Term	*a* Observed value, K	*b* Calculated value, K	Diff	*c* Calculated value, K	Diff
$(3s)^5S$	73 768	73 766	−2	73 702	−66
$(4s)^5S$	95 476	95 433	+33	95 444	−32
$(3s)^3S$	76 795	76 544	−251	76 798	+3
$(4s)^3S$	96 226	96 513	+287	96 324	+98
$(3s)^3D$	101 143	101 130	−13	101 168	+25
$(4s)^3D$	[122 250]	122 335	+85	122 282	+32
$(3s)^1D$	102 662	102 511	−151	102 718	+56
$(4s)^1D$	122 799	122 878	+79	122 720	−79
$(3s)^3P$	113 916	110 677	−3239	113 715	−201
$(4s)^3P$	[135 733]	140 327	+4590	135 858	+125
$(sp^5)^3P$	126 304	124 952	−1352	126 380	+76
$(3s)^1P$	115 918	119 470	+3552	115 859	−59
$(4s)^1P$	[136 083]	140 515	+4432	135 904	−179
$(sp^5)^1P$	[189 837]	181 855	−7982	190 077	+240
		$(D^2/N)^{\frac{1}{2}}$= 3042		$(D^2/N)^{\frac{1}{2}}$= 105	
Values of parameters used:					
$E_0(3s)$		120 385 K		114 880 K	
$E_0(4s)$		140 534		134 798	
$E_0(sp^5)$		180 921		189 902	
F_2		2961		2585	
$G_1(3s)$		730		796	
$G_1(4s)$		232		198	
$G_1(sp^5)$		31 982		31 949	
R_0		−1485		917	
R_1		−626		499	
R_1'		5075		2061	
R_1''		−1286		1249	
a		...		565	

the standard deviation to 1463 K. This particular procedure is described, and the term values obtained with it are reported, so that they may be compared with the term values obtained by a modification of this procedure which is discussed in the next paragraph.

In order to improve the fit, the $L(L+1)$ correction of Trees[14] was included in the diagonal elements of the $2p^33s$ and $2p^34s$ configurations with α taken to be the same for both configurations. This correction was not included in the diagonal terms of $2s2p^5$ since any such correction could be absorbed in the value of $E_0(sp^5)$. The same procedure was followed as that described above, and the value of α was determined by least squares using the diagonal sums. The results of this calculation are given in column *c* of Table II. The agreement between calculated and experimental values is all that can be expected in view of the extrapolation required for some of the experimental values. The extrapolated values are enclosed in brackets in column *a* of Table II. The fit might possibly be improved slightly by the inclusion of the three-by-three matrices in the determination of the parameters which effect the separation of the terms of the submatrices, but because of the other sources of error present this does not seem to be justified.

EVEN-PARITY CONFIGURATIONS

The only configuration which was included in the interaction with the $2p^33p$ configuration was the one with terms lying directly above it, $2p^34p$. There are, for these configurations, off-diagonal elements connecting terms having the same term designation in the same configuration but based on different terms of the parent ion, in addition to the off-diagonal elements connecting terms in different configurations.

The wave functions for the p^2p configurations were determined in the same manner as were the p^3s wave functions, by the application of the vector coupling formulas given in TAS 6[8]. With these wave functions, the matrix elements for electrostatic interaction were calculated, and the results, given in Table III, were compared with the matrix elements determined by Yamanouchi.[15] They are in accord except for the sign of the coefficient of G_2 in the matrix element connecting $(^2D)^3P$ with $(^2P)^3P$ and the signs of four of the off-diagonal elements connecting terms with different parentage. Yamanouchi obtained these off-diagonal elements by the method of spectroscopic stability, and the relative phases of his wave functions cannot be determined. In this paper, the relative phases of the wave functions for the terms of the parent ion were made to agree with Racah.[16]

A preliminary calculation was made on the $2p^33p$ configuration, ignoring the off-diagonal matrix elements, in order to obtain approximate magnitudes of the

TABLE III. Matrix elements of even-parity configurations.

Term	Diagonal elements[a]	Off-diagonal elements, different configurations, same parentage
$(^3D)^3F$	$E_0-6F_2-G_0-4G_2$	$R_0-R_0'-4R_1'$
$(^3D)^1F$	$E_0-6F_2-G_0+2G_2$	$R_0-R_0'+2R_1'$
$(^3D)^3D$	$E_0-6F_2-G_0-(17/2)G_2$	$R_0-R_0'-(17/2)R_1'$
$(^3P)^3D$	$E_0-G_0-(11/2)G_2$	$R_0-R_0'-(11/2)R_1'$
$(^1D)^1D$	$E_0-6F_2+2G_0+\frac{1}{2}G_2$	$R_0+2R_0'+\frac{1}{2}R_1'$
$(^3P)^1D$	$E_0+(3/2)G_2$	$R_0+(3/2)R_1'$
$(^5S)^5P$	$E_0-15F_2-G_0-10G_2$	$R_0-R_0'-10R_1'$
$(^3S)^3P$	$E_0-15F_2+(1/3)G_0+(10/3)G_2$	$R_0+(1/3)R_0'+(10/3)R_1'$
$(^3D)^3P$	$E_0-6F_2+(2/3)G_0-(59/6)G_2$	$R_0+(2/3)R_0'-(59/6)R_1'$
$(^3P)^3P$	$E_0-(15/2)G_2$	$R_0-(15/2)R_1'$
$(^3D)^1P$	$E_0-6F_2-G_0-(11/2)G_2$	$R_0-R_0'-(11/2)R_1'$
$(^3P)^1P$	$E_0-G_0-(5/2)G_2$	$R_0-R_0'-(5/2)R_1'$
$(^3P)^5S$	$E_0-G_0-10G_2$	$R_0-R_0'-10R_1'$
$(^3P)^3S$	E_0+3G_0	R_0+3R_0'

Connected terms	Off-diagonal elements, same configuration, different parentage	Off-diagonal elements, different configurations, different parentage
$(^3D)^3D-(^3P)^3D$	$\frac{1}{2}\sqrt{3}(-2F_2'+G_2)$	$\frac{1}{2}\sqrt{3}(-2R_1+R_1')$
$(^3D)^1D-(^3P)^1D$	$\sqrt{3}[-3F_2'+G_0-(7/2)G_2]$	$\sqrt{3}[-3R_1+R_0'-(7/2)R_1']$
$(^3D)^3P-(^3P)^3P$	$(\frac{1}{2}\sqrt{15})(9F_2'-G_0+\frac{1}{2}G_2)$	$(\frac{1}{2}\sqrt{15})(9R_1-R_0'+\frac{1}{2}R_1')$
$(^3D)^3P-(^5S)^3P$	$-(\frac{1}{2}\sqrt{5})(G_0+G_2)$	$-(\frac{1}{2}\sqrt{5})(R_0'+R_1')$
$(^3P)^3P-(^5S)^3P$	$\frac{1}{2}\sqrt{3}(G_0-5G_2)$	$\frac{1}{2}\sqrt{3}(R_0'-5R_1')$
$(^3D)^1P-(^3P)^1P$	$(\frac{1}{2}\sqrt{15})(2F_2'-G_2)$	$(\frac{1}{2}\sqrt{15})(2R_1-R_1')$

$E_0=3F_0(2p,2p)+3F_0(2p,np)+$integrals over closed shells,
$R_0=3R_0(2p3p,2p4p)+$integrals over closed shells,

$F_2=F_2(2p,2p)$ $\quad R_0'=R_0(2p3p,4p2p)$,
$F_2'=F_2(2p,np)$ $\quad R_1=R_1(2p3p,2p4p)$,
$G_0=G_0(2p,np)$ $\quad R_1'=R_1(2p3p,4p2p)$,
$G_2=G_2(2p,np)$.

[a] Diagonal elements of the two configurations have the same form and are obtained by replacing either $3p$, or $4p$ for np in the parameters.

[14] R. E. Trees, Phys. Rev. **84**, 1089 (1951).

[15] T. Yamanouchi, Proc. Phys.-Math. Soc. Japan **20**, 547 (1938).

[16] G. Racah, Phys. Rev. **63**, 367 (1943).

parameters G_0 and G_2. The $L(L+1)$ correction of Trees was included with the L value from the terms in p^3p. The value of F_2 was taken to be the same as for the $2p^3ns$ configurations. When a least squares calculation was made to determine the values of E_0, G_0, G_2, and α, it was discovered that the terms grouped themselves in such a way that the terms having the same parentage had approximately the same error. Other instances where the errors in the calculated term values are a function of the L value of the parent term have been reported by Racah.[17] If the errors in the terms built on the (^4S) parent are minimized by adjusting E_0, the errors in the other terms were approximately equal to the $L(L+1)$ correction to the $2p^3ns$ configurations for terms with the same parents. This suggested that the $L(L+1)$ correction, in all the configurations of the form al^nbl', was that of the parent ion.

To check this suggestion, the values of F_2 and α were calculated for the ground state of O II. This could be done exactly since there are three terms and three parameters. This calculation gave $F_2=2627$ K and $\alpha=531$ K. These values are in reasonable agreement with $F_2=2585$ K and $\alpha=565$ K found by least squares on the odd-parity configurations.

In addition Greyber[18] has shown that, if an interaction of the form $\sum_{a,b}\mathbf{l}_a\cdot\mathbf{l}_b$ is considered, for configurations of the form al^nbl' (which is the form considered here) the predominant part of the interaction energy will be proportional to $L(L+1)$ for the L of the parent term. In the configurations under consideration, therefore, the $L(L+1)$ correction is assigned to the parent ion. (This is equivalent to building the term values on the experimental values of the parent ion.)

TABLE IV. Term values (in K) of even-parity configurations.

Term	Observed	$2p^33p$ Calculated	Diff	Observed	$2p^34p$ Calculated	Diff
$(^4S)^5P$	86 629	86 393	−236	99 094	99 389	+295
$(^4S)^3P$	88 631	89 148	+517	99 680	102 810	+3130
$(^2D)^3D$	113 295	113 158	−137	125 780	125 932	+152
$(^2D)^3F$	113 719	113 698	−21		126 522	
$(^2D)^3P$		117 158			134 072	
$(^2D)^1P$		113 307			125 707	
$(^2D)^1F$	113 995	114 206	+211		127 042	
$(^2D)^1D$	116 630	117 294	+664		136 993	
$(^2P)^3S$		126 298			139 294	
$(^2P)^3D$	127 286	127 048	−238		139 672	
$(^2P)^3P$		127 811			145 782	
$(^2P)^1P$	127 667	127 940	+273		139 983	
$(^2P)^1D$	128 595	128 161	−434		149 991	
$(^2P)^1S$	130 943	130 886	−57		160 202	

Values of the parameters used, in K:

$E_0(3p)=$	131 233	$F_2'(3p)=$	258,
$E_0(4p)=$	139 883	$F_2'(4p)=$	81,
$G_0(3p)=$	4233	R_0	=5950,
$G_0(4p)=$	1671	R_0'	=2900,
$G_2(3p)=$	143	R_2	= 80,
$G_2(4p)=$	45	R_2'	= 108,
F_2 =	2585	α	= 565.

[17] G. Racah, Phys. Rev. 62, 438 (1942).
[18] H. Greyber, Ph.D. thesis, University of Pennsylvania, 1953 (unpublished).

TABLE V. Calculated values of σ^3.

Transition	Wave number of line	$\frac{a}{\sigma^3}$ Conf. int.	$\frac{b}{\sigma^3}$ Bates and Damgaard
$(^4S)3p\,^5P-(^4S)3s\,^5S$	12 861	9.85	7.86
$(^4S)3p\,^3P-(^4S)3s\,^3S$	11 836	10.75	9.53
$(^2D)3p\,^3F-(^2D)3s\,^3D$	12 576	10.20	8.80
$(^2D)3p\,^3D-(^2D)3s\,^3D$	12 152	10.38	8.62
$(^2D)3p\,^3P-(^2D)3s\,^3D$	[15 990]	5.98	
$(^2D)3p\,^1F-(^2D)3s\,^1D$	11 333	11.49	8.30
$(^2D)3p\,^1D-(^2D)3s\,^1D$	13 968	7.44	7.89
$(^2D)3p\,^1P-(^2D)3s\,^1D$	[10 589]	11.61	
$(^2P)3p\,^3D-(^2P)3s\,^3P$	13 370	7.31	7.63
$(^2P)3p\,^3P-(^2P)3s\,^3P$	[14 096]	6.76	
$(^2P)3p\,^3S-(^2P)3s\,^3P$	[12 583]	9.37	
$(^2P)3p\,^1D-(^2P)3s\,^1P$	12 677	9.57	8.41
$(^2P)3p\,^1P-(^2P)3s\,^1P$	11 749	8.24	8.48
$(^2P)3p\,^1S-(^2P)3s\,^1P$	15 025	6.28	7.64

In order to determine the radial integral parameters with configuration interaction included, some other procedure than the one applied in the case of the odd-parity configurations was necessary. This is so since there are only thirteen experimentally known term values, and there are in all twelve parameters to be determined, while the diagonal sum of but a single submatrix is known. To employ a least-squares calculation to determine all the parameters would be unsatisfactory. In a least-squares calculation no distinction is made between large and small parameters and in order to get the best possible agreement with experiment the calculation may give unreasonable magnitudes to some of the smaller parameters.

In order to calculate some of the parameters approximately, without resort to least squares, all those which consist of only one radial integral were calculated using the same radial wave functions employed in the calculation of the σ^3 for the oscillator strengths. This left $E_0(3p)$, $E_0(4p)$, and R_0 to be determined. They were determined by a least-squares calculation on all the two-by-two submatrices. This calculation resulted in values of $E_0(3p)$ and $E_0(4p)$ which made the diagonal element for $(4p)^1S$ slightly smaller than that for $(3p)^1S$. Some arbitrary adjustments were made in the larger parameters, therefore, in order to put these matrix elements in their proper order. This adjustment actually resulted in a better fit for all of the calculated term values. The term values obtained from the final calculations are given in Table IV.

Except for the values for the ^1P terms, the calculated energies are in reasonable agreement with the experimental values. The reason that the $(^4S)^3P$ terms do not agree with experiment seems to be that, with comparatively large off-diagonal elements (~6000 K to ~2000 K), configuration interaction with $2p^35p$ and $2p^36p$ would be important. The term $(^4S)5p\,^1P$ lies at 103 869 K and $(^4S)6p\,^1P$ lies at 105 911 K, about 4000 K and 6000 K respectively above the experimental value of the $(^4S)4p\,^1P$ term. Interaction with these configurations would lower the calculated term

TABLE VI. Absorption oscillator strengths $2p^3 3s - 2p^3 3p$.

Parent	$2p^3 3s$ level	$2p^3 3p$ level	f	Parent	$2p^3 3s$ level	$2p^3 3p$ level	f
(^3D)	$^3D_3$	$^3F_4$	0.448	(^3P)	$^3P_2$	$^3D_3$	0.219
	$^3D_3$	$^3F_3$	0.038		$^3P_2$	$^3D_2$	0.039
	$^3D_3$	$^3F_2$	0.001		$^3P_2$	$^3D_1$	0.003
	$^3D_2$	$^3F_3$	0.433		$^3P_1$	$^3D_2$	0.196
	$^3D_2$	$^3F_2$	0.048		$^3P_1$	$^3D_1$	0.065
	$^3D_1$	$^3F_2$	0.488		$^3P_0$	$^3D_1$	0.261
(^3D)	$^3D_3$	$^3D_3$	0.302	(^3P)	$^3P_2$	$^3P_2$	0.262
	$^3D_3$	$^3D_2$	0.038		$^3P_2$	$^3P_1$	0.087
	$^3D_2$	$^3D_3$	0.053		$^3P_1$	$^3P_2$	0.145
	$^3D_2$	$^3D_2$	0.237		$^3P_1$	$^3P_1$	0.087
	$^3D_2$	$^3D_1$	0.051		$^3P_1$	$^3P_0$	0.116
	$^3D_1$	$^3D_2$	0.085		$^3P_0$	$^3P_1$	0.349
	$^3D_1$	$^3D_1$	0.255	(^3P)	$^3P_2$	$^3S_1$	0.107
(^3D)	$^3D_3$	$^3P_2$	0.157		$^3P_1$	$^3S_1$	0.107
	$^3D_2$	$^3P_2$	0.039		$^3P_0$	$^3S_1$	0.107
	$^3D_2$	$^3P_1$	0.118	(^3P)	$^1P_1$	$^1D_2$	0.537
	$^3D_1$	$^3P_2$	0.003	(^3P)	$^1P_1$	$^1P_1$	0.272
	$^3D_1$	$^3P_1$	0.065	(^3P)	$^1P_1$	$^1S_0$	0.086
	$^3D_1$	$^3P_0$	0.087	(^5S)	$^5S_2$	$^5P_3$	0.479
(^3D)	$^1D_2$	$^1F_3$	0.470		$^5S_2$	$^5P_2$	0.342
(^3D)	$^1D_2$	$^1D_2$	0.296		$^5S_2$	$^5P_1$	0.205
(^3D)	$^1D_2$	$^1P_1$	0.202	(^5S)	$^3S_1$	$^3P_2$	0.605
					$^3S_1$	$^3P_1$	0.363
					$^3S_1$	$^3P_0$	0.121

values and would have the greater effect on the $(^4S)4p\,^3P$ term which has the larger error.

It should be pointed out that for configuration interaction between configurations of the form $al^n bl'$ the coefficients of the parameters in the off-diagonal elements are the same as those of corresponding parameters in the diagonal elements. Because of this, large diagonal elements will be associated with large off-diagonal elements. The resulting term values will be insensitive to changes in the parameters in the off-diagonal elements, since such changes can be approximately compensated by changes in the corresponding diagonal elements. This compensation will occur in each of the submatrices. As an example, where the coefficient of G_0 is positive that of R_0' will also be positive. As a result, making R_0' smaller and simultaneously making G_0 for the lower configuration larger and G_0 for the higher configuration smaller may result in little change in the term values. To see the effect of this, one can compare the values of G_0 and G_2 obtained by ignoring off-diagonal elements with those where G_0 and G_2 are first approximated by radial integrations and configuration interaction is included. Several investigators have determined these parameters without taking into account off-diagonal elements. Edlén[18] obtained $G_0=839$ K and $G_2=67$ K; Yamanouchi[15] did not give values for O I but they may be approximated to be $G_0\sim1000$ K and $G_2\sim50$ K by interpolation in the graphs in his paper. The preliminary calculations for this work gave $G_0=990$ K and $G_2=46$ K. In the final calculations including configuration interaction the values $G_0=4233$ K and $G_2=143$ K were used.

RESULTS

When the term values have been determined it is a straightforward calculation to determine the transformation matrices which will diagonalize the individual submatrices and thus determine the a_1's and a_2's of Eq. (6). When this had been done, the various values of σ^2 were calculated from Eq. (8). It was necessary to evaluate only the four integrals between the two sets of configurations $2p^3ns$ and $2p^3np$[19] since the integrals between $2p^3np$ and $2s2p^5$ will vanish. The values for σ^2 thus calculated are given in column *a* of Table V. In column *b* of Table V are given values of σ^2 calculated by the method of Bates and Damgaard[2] which uses hydrogenic wave functions and effective quantum numbers determined from experimental energies. The values for the wave numbers given in Table V are the differences between the centers of gravity of the experimental term values where these are known. If the line has not been identified, the values given (enclosed in brackets) are the differences between the calculated term values. Table VI gives the absorption oscillator strengths calculated by Eqs. (3) and (1).

Except for the transitions arising from ^1P levels, the values of σ^2 are as accurate as one might expect using the one-electron approximation and Hartree radial wave functions without exchange. The Hartree wave functions were determined by a self-consistent field calculation for the $2p^3 3s$ configuration. The wave functions for $3p$, $4s$, and $4p$ electrons were calculated assuming the wave functions of the core electrons to be the same as for $2p^3 3s$.

The results of the calculations could be improved if more energy levels were known for $2p^3 4p$ and if higher even-parity configurations were included in the calculations. This latter is not possible at present since only one term value is known for each of the next two configurations.

In conclusion, the author wishes to express his appreciation to Professor C. W. Ufford for suggesting this problem and for the valuable and patient assistance which he has given the author throughout the calculations. He also wishes to express his thanks to Professor L. C. Green for the help received during several informative discussions.

[19] The calculated values of the dipole velocity integrals are -5.812 for $3s-3p$, -0.504 for $3s-4p$, 5.547 for $4s-3p$, and -12.395 for $4s-4p$. As a check on the wave functions, the dipole moment integral was calculated for $3s-3p$. It was -5.818.

Line Broadening by Electrons: Validity of Simple Theories*

ROLAND E. MEYEROTT, *The RAND Corporation, Santa Monica, California*

AND

HENRY MARGENAU, *Yale University, New Haven, Connecticut*

(Received November 29, 1955; revised manuscript received June 6, 1955)

A comparison is made between simple impact theories of line broadening by electrons (Weisskopf, Lindholm, Inglis and Teller, and Unsöld) and recent more detailed quantum calculations by Kivel, Bloom, and Margenau.

THE difficulties of a rigorous treatment of molecular disturbances upon the radiative process have forced into use a variety of simple schemes for dealing with problems of pressure broadening. Impact theories, statistical theory and combinations thereof, enjoy applications that are sometimes hard to justify. Particularly troublesome are the effects of electron collisions upon the shape of spectral lines because here, in addition to the normal difficulties, one faces the problem as to the extent to which a perturbing electron can be localized during its action upon a radiating atom, as well as the question of the appropriate potential to be employed.

Simplified theories dealing with the electron problem[1,2] ingeneously use the physical ideas of impacts and the notions of Lorentz and Weisskopf to arrive at answers for line widths. The answers are not easily tested experimentally and rely for their credibility upon consistency with other findings. In a preceding paper by Kivel, Bloom, and Margenau,[3] a detailed quantum mechanical theory of the electron effect has been developed for a case (the Lyman Alpha line) which, while holding little intrinsic interest, permits the qualitative theories to be tested. This was indeed the chief motive for the work. The present paper makes a comparison of the more complete theory with various impact theories and with an analysis by Inglis and Teller.[2]

I.

In Secs. I to IV of this article attention will be confined to the effect called "universal" in K.B.M. It accounts for the line width in the absence of degeneracy and was computed by supposing that the atom has only two levels, the upper and the lower of the radiative transition. The possibility of transitions to higher states was disregarded because their contributions to the line width was found to be small *in comparison with the linear Stark effect* (arising from the presence of degeneracy). For this reason, as we shall see, the universal effect as defined in K.B.M. does not include completely what is sometimes called the quadratic Stark effect. This should be emphasized because a statement in the abstract of K.B.M. is misleading.

In that paper the line width arising from the cause in question was found to be

$$\gamma_U = \nu_c, \tag{1}$$

but ν_c, the collision frequency, instead of being related to the ordinary collision cross section $\sigma' = \pi\rho^2$ by the equation $\nu_c = n v \sigma'$, is now defined through the de Broglie cross section $\sigma = \pi(\hbar/mv)^2$ and a numerical "efficiency factor" $2S_U$ as follows:

$$\nu_c = n v \sigma \cdot 2S_U. \tag{2}$$

The function S_U was plotted in reference 3; roughly, it is approximated by

$$S_U = 4.3(1-e^{-4.1x}), \tag{3}$$

if x is the energy of the perturbing electron in units of $e^2/2a$, where e is the electronic charge and a is the first Bohr radius. A plot of γ_U, based on (2) and (3), is found in Fig. 1.

II.

Our first comparison will be with a simple form of impact theory. For details, we refer to a review by Unsöld.[1] This theory assumes that an electron, located at a distance r from the radiating atom, produces a line shift of circular frequency

$$\Delta\omega = (e/\hbar)V(r), \tag{4}$$

where $V(r)$ is the potential due to the atomic charge distribution at point r. This line shift is computed for an electron passage with closest distance of approach ρ at constant electron velocity, v, and all those passages are declared effective in broadening the line for which the phase change, $\eta = \int \Delta\omega dt$ is greater than some number η_0, of order 1. Thus is defined an optical collision diameter, ρ_0, which is made to function in the line width formula as does Lorentz' kinetic theory diameter in that author's well-known theory. The collision frequency is then given as

$$\gamma_c = n v \pi \rho_0^2, \tag{5}$$

* This work was sponsored by the U. S. Atomic Energy Commission.

[1] A. Unsöld, Vierteljahrsschr. Astron. Ges. **78**, 213 (1943).

[2] D. R. Inglis and E. Teller, Astrophys. J. **90**, 439 (1939).

[3] Kivel, Bloom, and Margenau, Phys. Rev. **98**, 495 (1955) (hereafter referred to as K.B.M.).

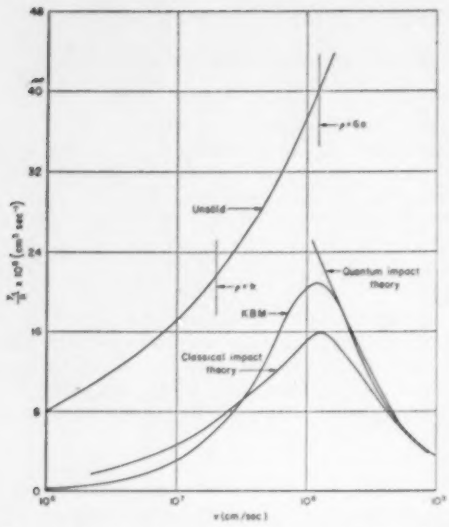


FIG. 1. γ_c/n plotted as a function of velocity. The classical impact theory curve was computed by the method of Sec. II while the quantum impact theory was computed by the method of Sec. III.

where n is the electron density, and the radiation intensity is

$$I(\omega)=\text{const}\,(\omega^2+\gamma_c^2)^{-1}, \tag{6}$$

where ω is the angular frequency measured from the line center.

This simplest form of the impact theory, proposed by Weisskopf, has often been employed in the older literature, and we shall give an example of it in Sec. IV. At this point, we adopt an improvement by Lindholm[4] which provides a method for computing the value of the limiting phase, η_0, to be used in determining ρ_0. According to Lindholm, γ_c in Eq. (6) is approximated by

$$\gamma_c=nv\sum_l\sigma_l[1-\cos(\eta_l-\eta_l')], \tag{7}$$

where, in a quantum calculation of the collision process' η_l is the phase shift suffered by an electron in the lth angular momentum state as it interacts with the excited atom, while η_l' is the same quantity for the atom in the ground state.

Furthermore, $\sigma_l=(2l+1)\sigma$, and σ is the de Broglie cross section $\pi(\hbar/mv)^2$. The classical approximation to (7) is

$$\gamma_c=nv\int_0^\infty 4\pi\rho d\rho\,\sin^2\eta/2. \tag{8}$$

In this expression η is the phase change computed from Eq. (4).

In K.B.M., the potential energy between electron and atom entered in its classical form, containing the coordinates of the atomic as well as of the perturbing electron. For the calculation of η a suitable average no longer involving the coordinates of the atomic electron must be chosen. This is the potential of the atomic distribution, averaged over angles, that acts on the external electron. Thus

$$V(r)=V_{2p}(r)-V_{1s}(r),$$

where

$$V_{2p}(r)=-\frac{e}{a}\left[\frac{1}{r}+\frac{3}{4}+\frac{r}{4}+\frac{r^2}{24}\right]e^{-r} \tag{9}$$

and

$$V_{1s}(r)=-\frac{e}{a}\left[\frac{1}{r}+1\right]e^{-2r}, \tag{10}$$

and r is in units a, the first Bohr radius. The replacement of V by the difference between the two states, though not always made, is necessary for comparison with K.B.M. and has an obvious justification. We then take

$$\eta=\frac{2}{v}f(\rho)$$

$$=-\frac{2}{v}\int_0^\infty[V_{2p}((\rho^2+x^2)^{\frac{1}{2}})-V_{1s}((\rho^2+x^2)^{\frac{1}{2}})]dx. \tag{11}$$

The integrations in Eqs. (11) and (8) are carried out numerically and the results are shown in Fig. 1 where γ_c/n is plotted as a function of v. The average impact parameter ρ_0 effective for broadening can be obtained from Eq. (5). In order to permit an estimate of the importance of quantum effects for this problem the quantity ρ_0/λ where $\lambda=(\hbar/mv)$ has also been plotted in Fig. 2.

The agreement between the quantum mechanical calculation and the classical one is good over the range where $(\rho_0/\lambda)>1$, the maximum deviation in this region being about 25%.

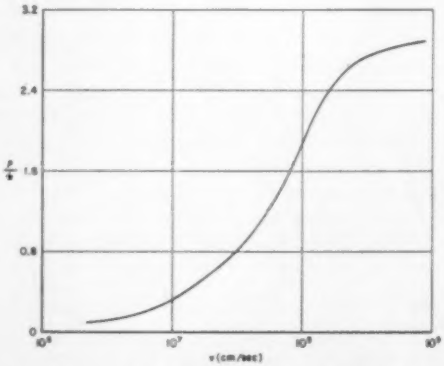


FIG. 2. ρ/λ for classical impact theory plotted as a function of velocity.

[4] E. Lindholm, Archiv Mat. Astron. Fysik. **B28**, No. 3 (1941).

III.

A further comparison can be made by evaluating Eq. (7), using the phase shifts computed for $V_{2p}(r)$ and $V_{1s}(r)$ in Eqs. (9) and (10) respectively. The phase-shifts for $V_{2p}(r)$ have been computed by Swan[5] while those for $l=0$ for $V_{1s}(r)$ are given by Massey and Moiseiwitsch.[6] Since the phase shifts for $l>0$ are not negligible for $V_{1s}(r)$, the additional ones were computed by Langer's[7] method and are given in Table I for reference. These values of the phase shifts along with those of Swan for $2p$ have been used to evaluate ν_v in Eq. (7) and the results are also shown in Fig. 1. All the points except that for $v=1.1\times10^8$ cm/sec fall on the K.B.M. curve. This seems somewhat remarkable in view of the fact that the elastic scattering cross section obtained by Swan upon using the phase shifts computed according to Langer's method and the results of the Born approximation differ by a factor of two in this velocity range. Considering the fact that K.B.M. in effect used the Born approximation it would seem that the broadening must be less sensitive to the method of calculation than is the elastic scattering cross section.

IV.

In dealing with solar Mg-lines—which are known to undergo only quadratic Stark effect—Unsöld[8] employs the impact theory in the following way. Assuming a static Stark effect potential and neglecting inhomogeneity of the field he takes $\Delta\omega$ to be C/r^4.

Computation then leads to the result for the half-width:

$$\gamma_c=\pi^{7/3}C^{2/3}v^{1/3}n \tag{12}$$

in terms of variables previously defined and n, the number of electrons per cm^3. The constant C for the L_α-line is simply related to the second-order depressions of the two-quantum levels and the ground level of hydrogen in an electric field. These are given by $\Delta E=-\alpha F^2$ if F is the field strength. For the three two-quantum states which combine with the ground state α has the values 84, 78, 78 in atomic units, or a mean of 80. For the ground state, $\alpha=9/4$. This gives $\Delta\omega=77.75(a^3F^2/\hbar)$, a being the first Bohr radius. When F is considered to be ϵ/r^2 we find

$$C=4.1\times10^{-16}\text{ cm}^4\text{ sec}^{-1},$$

TABLE I. Scattering phase shifts, $\eta_l(1s)$, from V_{1s} potential.

k	v(cm sec^{-1})	$\eta_0(1s)$	$\eta_1(1s)$	$\eta_2(1s)$	$\eta_3(1s)$	$\eta_4(1s)$
0.500	1.10×10^8	1.044	0.0099	...	...	...
0.831	1.82×10^8	0.952	0.0701	0.00208	...	...
1.803	4.02×10^8	0.726	0.218	0.0616	0.0146	...
2.7187	5.97×10^8	0.590	0.252	0.110	0.0533	0.024

[5] P. Swan, Proc. Phys. Soc. (London) **A67**, 1086 (1954).
[6] H. S. W. Massey and B. L. Moiseiwitsch, Proc. Roy. Soc. (London) **A205**, 483 (1951).
[7] R. Langer, Phys. Rev. **51**, 669 (1937), see also reference 5.
[8] A. Unsöld, Z. Astrophysik **12**, 56 (1936); **23**, 75 (1944).

the ratio γ_c/n from Eq. (5), with this value of C, is plotted in Fig. 1. While Unsöld's curve is everywhere proportional to $v^{\frac{1}{3}}$, that of K.B.M. varies as v for small, as v^{-1} for high velocities. In this comparison we have taken η_0 to be 1. A value of 0.64, which is indicated by Lindholm's reasoning, would raise the impact curve and heighten the disagreement.

In addition to the quantum limitation to the impact parameter, $\rho_0>(\hbar/mv)$, mentioned before, this treatment becomes invalid if $\rho_0<6a_0$. For smaller impact parameters the perturber penetrates the charge distribution of the atom and classically the potential should go over smoothly to that of Eqs. (9) and (10). A third limit which is of little consequence at ordinary density is $\rho_0<n^{-\frac{1}{3}}$, requiring that only one perturbing electron counts at one time. The first two limits are indicated in Fig. 1. (It should be remembered that $\rho_0=(\hbar/mv)$ for the lower limit in the Unsöld curve is computed from a different potential than ρ_0/λ plotted in Fig. 2. Thus $\rho_0=\lambda$ is satisfied for a different velocity in the two cases.)

As already noted, this curve cannot be compared directly with K.B.M. To see whether it represents an additional effect of greater magnitude than the universal one an extension of the latter work to include transitions to excited states is required. This calculation is in progress.

V.

A simple impact theory of line broadening appropriate when a linear Stark effect occurs is given by Inglis and Teller.[2] For fast collisions, such that $v>(1/\sqrt{2})v_0$, where v_0 is the velocity of the electron in the first Bohr orbit (in the case of Lyman Alpha line) they obtain the following expression for the broadening

$$\gamma=16(nv\sigma)[1+2\log\rho_m/\rho_1]. \tag{13}$$

Here σ is again the de Broglie cross section, ρ_m is the largest unscreened impact radius, $(\pi n)^{-\frac{1}{3}}$, and ρ_1 is the largest impact radius that causes a transition in the atom with certainty, $4a_0v_0/v$.

This result should be comparable with γ_p of reference (3). The expression given for γ_p in that paper is

$$\gamma_p=2nv\sigma\{\tfrac{1}{6}(S_p)_{2p_0,2s_0}+(6/35)[1-(1+x_+)^{-7}]\}. \tag{14}$$

When $x_+>\bar{x}$, S_p becomes

$$(S_p)_{2p_0,2s_0}=9[\tfrac{1}{2}+\log(x_+/\bar{x})]. \tag{15}$$

x_+ and $\bar{x}$ are related to ρ_m and ρ_1 by the following expression:

$$(x_+/\bar{x})^{\frac{1}{2}}=4(\rho_m/\rho_1),$$

and $x_+\gg1$. Hence, γ_p becomes essentially

$$\gamma_p=3nv\sigma[3.6+2\log(\rho_m/\rho_1)]. \tag{16}$$

The factor 16 in the Eq. (13) is essentially the square of the linear Stark coefficient in the proper units. If the correct value 9 is taken for this quantity instead of

16 as used in Eq. (13) the agreement between the qualitative result of Inglis and Teller and that of K.B.M. is improved. For most cases of practical interest $\log(\rho_m/\rho_2)\sim 10$ so that the order of magnitude of the agreement is certainly interesting for densities of practical interest.

For velocities less than $v_0/\sqrt{2}$, Inglis and Teller include the electrons with the ions and treat the entire effect according to the statistical theory. K.B.M. deal with *sharp* electron velocities; their formulas will therefore not reduce to those of the statistical theory, which can only result from a distribution of velocities that allows the electrons to be localized.[9] Hence, in this limit the theories cannot be compared.

[9] L. Spitzer, Jr. Phys. Rev. **55**, 699 (1939); H. Margenau and B. Kivel, Phys. Rev. **98**, 1822 (1955).

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Temperature Behavior of the Thomas-Fermi Statistical Model for Atoms

RICHARD LATTER
The Rand Corporation, Santa Monica, California
(Received May 2, 1955)

The usual theoretical calculations of equations of state and specific heats, particularly at high temperatures and pressures, are dependent on the Thomas-Fermi statistical model of the atom. The mathematical description of this model involves complicated nonlinear differential equations, for which there have been an inadequate number of solutions available in the past. A number of solutions sufficiently extensive to determine the thermodynamic properties of all elements over an exceedingly wide range of temperatures and densities have been obtained with the aid of an IBM 701 Defense Calculator. The results are presented in graphical form.

In addition, some of the analytic properties of the Thomas-Fermi equations have been investigated and certain approximate analytic solutions have been derived for limiting cases.

I. INTRODUCTION

THEORETICAL calculations with the Thomas-Fermi atomic model are limited by the necessity of solving a difficult nonlinear differential equation. Only rough analytic approximations have been developed in previous investigations and these are primarily asymptotic solutions with a very restricted range of applicability. Principal emphasis in recent years has been toward the numerical treatment of the model. However, the results available at the present time are almost exclusively concerned with the properties of the model at zero temperature. For the temperature behavior of the Thomas-Fermi model the most detailed calculations have been in the limit of small temperatures and small volumes where the problem may be treated as a perturbation from the zero-temperature model. The formulation and solution of the temperature-perturbed model have been given by Bethe and Marshak.[1] Numerical results have been obtained for this theory by Feynman, Metropolis and Teller,[2] and Gilvarry and Peebles.[3] The temperature perturbation theory has the disadvantage, however, that it requires considerable numerical work and, moreover, that it is limited only to the region of large positive work functions. The negative and small positive work functions encompass, on the other hand, an extensive region of interesting atomic states.

The exact treatment of the temperature effects, without the Bethe-Marshak approximation, does not appear to involve appreciably more numerical work than with the approximation, and, of course, it is not restricted to particular physical states. Unfortunately only a few isolated solutions have been obtained, namely by Feynman, Metropolis and Teller,[2] using the exact Thomas-Fermi model. The calculations to be reported here extend these previous solutions so as to provide a detailed and comprehensive description of the thermodynamic behavior of the model. Exchange effects (see Appendix I) were not included. But in determining the electronic specific heat it is known that exchange effects are almost completely cancelled by the effects of correlation at low temperatures[4]; at high temperatures exchange effects become negligible. And therefore the temperature behavior of the atom may be more adequately described by ignoring exchange altogether if the correlation is not included.

II. DESCRIPTION OF THE THOMAS-FERMI MODEL WITH TEMPERATURE

The statistical model of the atom treats the equilibrium properties of Z electrons constrained to lie within a sphere of radius r_0, at the center of which is a fixed positive charge, $+Ze$. The extension of the model to include ions consists in allowing more or fewer electrons within the sphere than positive charges at the center. The present discussion and calculations will be

[1] H. A. Bethe and R. E. Marshak, Astrophys. J. **91**, 239 (1940).
[2] Feynman, Metropolis, and Teller, Phys. Rev. **75**, 1561 (1949).
[3] J. J. Gilvarry and G. Peebles, Phys. Rev. **99**, 550 (1955).
[4] D. Pines, Phys. Rev. **92**, 626 (1953).

limited to the neutral atom; but ions may be treated by only minor changes in the following arguments.

The electrostatic potential is related to the atomic charge distribution through Poisson's equation[5]

$$\nabla^2 V(\mathbf{r}) = 4\pi e\rho(\mathbf{r}). \tag{1}$$

The Thomas-Fermi model provides a simple estimate of the charge density $-e\rho(\mathbf{r})$ in terms of the potential function $V(\mathbf{r})$. The arguments leading to this connection have been given from a number of independent points of view for the case of zero temperature.[6,7] The extension to nonzero temperatures was given by Feynman, Metropolis and Teller.[2] The result of this extension is that

$$\rho(\mathbf{r}) = \frac{8\pi}{h^3}\int_0^\infty p^2 dp \times \frac{1}{\exp\{[-\alpha + p^2/2m - eV(\mathbf{r})]/kT\}+1}. \tag{2}$$

Equations (1) and (2) can be simplified in appearance by making a few notational changes and requiring that the potential and charge distributions be spherically symmetric. Then it is found that

$$\phi''(x) = axI_{\frac{1}{2}}(\phi(x)/x), \tag{3}$$

where

$$\phi(x) = [\alpha + eV(r)]r/kTr_0, \quad x = r/r_0, \quad a = (r_0/c)^2,$$
$$1/c = 4\pi e(2m)^{\frac{1}{2}}(kT)^{\frac{1}{2}}/h^3,$$

and the Fermi-Dirac function is defined for general n by

$$I_n(\eta) = \int_0^\infty dy y^n \frac{1}{\exp(y-\eta)+1}. \tag{4}$$

The zero of the potential function is chosen at $r = r_0$. The differential equation, Eq. (3), must be supplemented with the boundary conditions on the function $\phi(x)$. These conditions state, on the one hand, that the atomic system is neutral so that the potential gradient at the boundary of the atom is zero and, on the other hand, that the potential function has a singularity at the center of the atom corresponding to a charge $+Ze$ at that point. These boundary conditions, when expressed in the present notation, become

$$\phi'(1) = \phi(1), \tag{5}$$

and

$$\phi(0) = Ze^2/kTr_0. \tag{6}$$

All the thermodynamic properties of the atomic system may be related to the solutions of these equations, Eqs. (3), (5), and (6). The quantities of particular interest are the pressure, kinetic and potential energies, and the entropy, since in terms of these quantities all other thermodynamic functions may be derived explicitly.

The pressure is expressed as the rate of transfer of momentum between the electrons and the atomic boundary. Since at the boundary of the atom the electric field is zero, and consequently the potential is a constant, the electrons behave locally as a uniform gas wherein the only interchange of electron momenta is with the boundary. According to Eq. (2) the density of electrons at position $\mathbf{r}$ with momenta in the range $d\mathbf{p}$ is

$$\rho(\mathbf{r},\mathbf{p}) = \frac{8\pi}{h^3}\frac{p^2}{\exp\{[-\alpha + p^2/2m - eV(r)]/kT\}+1}. \tag{7}$$

Since the electrons are moving isotropically, the rate of momentum transfer to the boundary per electron is

$$(2p)(p/3m), \tag{8}$$

and, hence, the pressure is

$$P = \frac{8\pi}{h^3}\int_0^\infty dp \frac{p^3}{\exp[(-\alpha + p^2/2m)/kT]+1} \times (2p)(p/3m), \tag{9}$$

where $V(r_0) = 0$ has been used. With the help of Eq. (4) this expression for the pressure reduces to

$$P = \frac{ZkT}{V}\frac{2}{9}\frac{a}{\phi(0)}I_{\frac{1}{2}}[\phi(1)]. \tag{10}$$

The kinetic energy per electron is $p^2/2m$, which multiplied by the density of electrons from Eq. (7) and integrated over all momenta and the volume of the atom, leads to the total kinetic energy of the atom. Thus

$$E_{\text{kin}} = \frac{32\pi^2}{h^3}\int_0^{r_0} r^2 dr \int_0^\infty dp \times \frac{p^2}{\exp\{[-\alpha + p^2/2m - eV(r)]/kT\}+1}\frac{p^2}{2m} \tag{11}$$

or, in the present notation,

$$E_{\text{kin}} = ZkT\frac{a}{\phi(0)}\int_0^1 dx x^2 I_{\frac{1}{2}}\left[\frac{\phi(x)}{x}\right]. \tag{12}$$

The potential energy of the electrons results from their mutual interaction and their interaction with the nuclear field. This energy may be expressed as

$$E_{\text{pot}} = -\tfrac{1}{2}(4\pi e)\int_0^{r_0} r^2 dr\rho(r)V_e(r) - 4\pi e\int_0^{r_0} r^2 dr\rho(r)V_n(r), \tag{13}$$

[5] Throughout this discussion e refers to the magnitude of the electronic charge.

[6] E. Fermi, Atti. Accad. Nazl. Lincei Rend. Classe Fis. Matt. E. Nat. **6**, 602 (1927); Z. Physik **48**, 73 (1928).

[7] L. Brillouin, Actualités Sci. et Ind. **160**, 1 (1934).

where $V_e(r)$ is the potential arising from the electrons and $V_n(r)$ the potential from the nuclear field. This may be written

$$E_{\text{pot}}=-\tfrac{1}{2}(4\pi\epsilon)\int_0^{r_0} r^2 dr\rho(r)\left[V(r)+\frac{Z}{r}\right]. \quad (14)$$

Using Eq. (2) and re-expressing Eq. (14) in terms of the present notation, it is found that

$$E_{\text{pot}}=-ZkT\frac{a}{2\phi(0)}\int_0^1 dx I_{\frac{1}{2}}\left[\frac{\phi(x)}{x}\right] \times x[\phi(x)-\phi(1)x+\phi(0)]. \quad (15)$$

An important relationship exists between the quantities E_{pot}, E_{kin}, and P, namely,

$$2E_{\text{kin}}+E_{\text{pot}}=3PV. \quad (16)$$

This expression may be derived directly from Eqs. (10), (12), and (15) or from the virial theorem, which states that

$$2E_{\text{kin}}=I. \quad (17)$$

The approach from the virial theorem is simpler and consists in the evaluation of the virial $I=-\langle\sum_i \mathbf{r}_i\cdot\mathbf{F}_i\rangle_{Av}$. The angular braces indicate a space (or time) average; the subscript i refers to the ith electron, which is subjected to the force $\mathbf{F}_i$ when at position $\mathbf{r}_i$. The force arises from the mutual electronic interaction, from the electron-nuclear interaction and from collisions with the atomic boundary. The virial I may be decomposed into terms arising from the three sources of the force $\mathbf{F}_i$ as follows:

$$I=I_{ee}+I_{en}+I_{eb}. \quad (18)$$

The last term describes the contribution from the atomic boundary and may be evaluated in terms of the applied pressure P. Thus, since the force applied by the boundary is directed toward the center of the sphere and occurs always at the atomic radius r_0,

$$I_{eb}=-\langle\textstyle\sum_i \mathbf{r}_i\cdot\mathbf{F}_{i(\text{boundary})}\rangle_{Av} =\langle r_0\textstyle\sum_i F_{i(\text{boundary})}\rangle_{Av}.$$

Now the time-average boundary force is just the pressure P times the area of the atomic surface and thus may be expressed as

$$I_{eb}=4\pi r_0^3 P=3PV, \quad (19)$$

where V is the volume of one atom. The second term in Eq. (18) results from the force exerted by the nucleus on the electrons. This force is directed toward the nucleus and has the magnitude Ze^2/r_i^2, where r_i is the radial distance to the ith electron. Thus

$$I_{en}=\left\langle\sum_i\frac{Ze^2}{r_i}\right\rangle_{Av},$$

and the space average of this sum is just the negative of the electron-nuclear potential energy, so that

$$I_{en}=-E_{\text{pot}}{}^{en}. \quad (20)$$

Finally, the term I_{ee} arises from mutual electronic forces. The force on the ith electron due to all other electrons may be written

$$\mathbf{F}_i=e^2\sum_j\frac{(\mathbf{r}_i-\mathbf{r}_j)}{|\mathbf{r}_i-\mathbf{r}_j|^3}.$$

Thus

$$I_{ee}=-e^2\left\langle\sum_i\sum_j\frac{\mathbf{r}_i\cdot(\mathbf{r}_i-\mathbf{r}_j)}{|\mathbf{r}_i-\mathbf{r}_j|^3}\right\rangle_{Av} =-\frac{e^2}{2}\left\langle\sum_i\sum_j\frac{(\mathbf{r}_i-\mathbf{r}_j)\cdot(\mathbf{r}_i-\mathbf{r}_j)}{|\mathbf{r}_i-\mathbf{r}_j|^3}\right\rangle_{Av} =-\frac{e^2}{2}\sum_i\sum_j\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|},$$

which is the negative of the mutual electron potential energy. And, hence,

$$I_{ee}=-E_{\text{pot}}{}^{ee}. \quad (21)$$

Since the total electronic potential energy is

$$E_{\text{pot}}=E_{\text{pot}}{}^{ee}+E_{\text{pot}}{}^{en},$$

Eqs. (17)–(21) establish the statement of Eq. (16).

The derivation of the entropy of the Thomas-Fermi model in terms of the solutions of Eq. (3) has been carried out by Brachman[8] and, more simply, by McMillan and Gilvarry.[9] An alternative derivation, presented below, starts from the partition function, which is related to the entropy through the equation

$$S=E/T+k\ln Q, \quad (22)$$

where E is the total internal energy of the atom and Q is the atomic partition function. It has not been possible so far to derive Q rigorously for the Thomas-Fermi model. A form for it which is, however, consistent with the assumptions of the model may be derived heuristically. Thus consider the atom divided into cells small enough so that the potential is essentially constant in each cell but large enough so that any electrons in a given cell occupy a region of phase space large compared to h^3. The Thomas-Fermi assumptions assert that the electrons occupy each cell independently of the occupation of the other cells within the atom and, moreover, obey Fermi-Dirac statistics. If now $n(\mathbf{r}_i,\mathbf{p},\sigma)$ denotes the number of electrons of momentum $\mathbf{p}$ and spin σ in the cell at position $\mathbf{r}_i$ (which number according to the Fermi-Dirac statistics is 0 or 1), then the total energy of the electrons in this cell is

$$\epsilon(\mathbf{r}_i)=\sum_{\mathbf{p},\sigma} n(\mathbf{r}_i,\mathbf{p},\sigma)[p^2/2m-eV(r_i)], \quad (23)$$

[8] M. Brachman, Phys. Rev. **84**, 1263 (1951).
[9] W. G. McMillan and J. J. Gilvarry (unpublished).

where $eV(\mathbf{r}_i)$ is the total potential energy of an electron in this cell. The total energy of the atom is not obtained simply by adding together the contributions $\epsilon(\mathbf{r}_i)$ from each cell since the potential energy E_{pot}^{ee} of mutual electronic interaction would be counted twice. Correcting for this duplication, the total energy of the atom becomes

$$E=-E_{\text{pot}}^{ee}+\sum_i \epsilon(\mathbf{r}_i). \tag{24}$$

The canonical partition function may now be written

$$Q=\sum_{\{n\}=Z} \exp[-[-E_{\text{pot}}^{ee}+\sum_i \epsilon(\mathbf{r}_i)]/kT], \tag{25}$$

where the first summation is over all values of $n(\mathbf{r}_i,\mathbf{p},\sigma)$ subject to the constraint that there are a total of Z electrons. Several approximations, which appear consistent with the model, must now be made in order to derive the Thomas-Fermi atom. First, E_{pot}^{ee} is treated as independent of the occupation numbers $n(\mathbf{r}_i,\mathbf{p},\sigma)$ and is taken equal to the mean potential energy of mutual interaction. This assumption is consistent with treating the potential function $V(\mathbf{r}_j)$ as independent of $n(\mathbf{r}_i,\mathbf{p},\sigma)$ for all i and j. If now $Z\alpha$, where α is the work-function determined so that Z is the total number of electrons, is added and subtracted in the exponent of Q, then Eq. (25) may be written with the aid of Eq. (23) as

$$Q=\sum_{\{n\}=Z} \exp\left\{-\left[-E_{\text{pot}}^{ee}+Z\alpha + \sum_{i,\mathbf{p},\sigma} n(\mathbf{r}_i,\mathbf{p},\sigma)\left(\frac{p^2}{2m}-eV(\mathbf{r}_i)-\alpha\right)\right]\Big/kT\right\}, \tag{26}$$

since $\sum_{i,\mathbf{p},\sigma} n(\mathbf{r}_i,\mathbf{p},\sigma)=Z$. The explicit evaluation of Q is very complicated if the constraint $\{n\}=Z$ is retained. However, if the number of electrons is large, this constraint is unimportant and may be removed. The unrestricted sum is easily evaluated due to the separability of the terms $n(\mathbf{r}_i,\mathbf{p},\sigma)$ in the exponent and is

$$Q=\exp\left[\frac{-(-E_{\text{pot}}^{ee}+Z\alpha)}{kT}\right] \times \prod_{i,\mathbf{p},\sigma}\left\{1+\exp\left[-\left(\frac{p^2}{2m}-eV(\mathbf{r}_i)-\alpha\right)\Big/kT\right]\right\}^2. \tag{27}$$

Taking the logarithm of both sides of this equation and replacing the summations by integrations lead to

$$\ln Q=-\frac{(-E_{\text{pot}}^{ee}+Z\alpha)}{kT}+\frac{32\pi^2}{h^3}\int_0^{r_0} r^2dr\int_0^\infty p^2dp \times \ln\left\{1+\exp\left[-\left(\frac{p^2}{2m}-eV(\mathbf{r}_i)-\alpha\right)\Big/kT\right]\right\}. \tag{28}$$

In terms of the previous notation, this equation becomes

$$\ln Q=\frac{E_{\text{pot}}^{ee}}{kT}-Z\phi(1)+\frac{a}{\phi(0)}\int_0^1 x^2dx\int_0^\infty q^{\frac{1}{2}}dq \times \ln\left\{1+\exp\left(-q+\frac{\phi(x)}{x}\right)\right\} \tag{29}$$

where $q=p^2/2mkT$. An integration by parts on q, using Eq. (4), gives finally

$$\ln Q=\frac{E_{\text{pot}}^{ee}}{kT}-Z\phi(1)+\frac{2}{3}\frac{a}{\phi(0)}\int_0^1 x^2dxI_{\frac{3}{2}}\left(\frac{\phi(x)}{x}\right). \tag{30}$$

It is easily shown that a functional differentiation of $\ln Q$ with respect to $\phi(x)/x$ leads to the Thomas-Fermi form for the charge density and hence Eq. (30) is, in this important respect, compatible with the model. Combining Eqs. (12), (30), and (22) yields the expression for the entropy

$$S/Zk=[(5/3)E_{\text{kin}}+2E_{\text{pot}}^{ee}+E_{\text{pot}}^{en}]/ZkT-\phi(1). \tag{31}$$

III. ANALYTIC PROPERTIES OF THE THOMAS-FERMI EQUATION

Such properties of Eq. (3) as are relevant to the numerical treatment below will now be investigated. For this purpose it is necessary to re-express Eq. (3) as an integral equation, which provides the basis for the numerical procedure to be used. The transformation of Eq. (3) to integral form consists of a straightforward double integration followed by an integration by parts and leads to the result

$$\phi(x)=\phi(1)x+a\int_x^1 dy(y-x)yI_{\frac{1}{2}}\left(\frac{\phi(y)}{y}\right), \tag{32}$$

which incorporates implicitly the boundary condition of Eq. (5). It is necessary, however, to impose an important restriction on the class of physically admissible solutions of Eq. (32), namely, that $\phi(x)$ be bounded on the closed interval $0\leq x\leq 1$. This restriction leads in turn to a considerable limitation on the range of admissible values for the parameters a and $\phi(1)$.

An explicit, analytic demarcation of the allowed region for which a and $\phi(1)$ lead to bounded solutions $\phi(x)$ has not been obtained, but the boundary curve has been found numerically in the region of negative $\phi(1)$ and by an approximate analytic solution for positive $\phi(1)$. If the boundary of the admissible region is denoted by the function $\phi(1)=f(a)$, the region is defined by

$$\phi(1)<f(a). \tag{33}$$

To determine the relation $\phi(1)=f(a)$, it is observed first that if

$$1\geq x_1>x_2\geq 0,$$

then Eq. (32), combined with the positivity of $I_{\frac{1}{2}}(\eta)$,

asserts that

$$0\leq\phi(x_1)-\phi(1)x_1<\phi(x_2)-\phi(1)x_2.$$

Thus, $\phi(x)-\phi(1)x$ is a positive, monotonic increasing function for decreasing x. If it is assumed that the solutions $\phi(x)$ of Eq. (32) depend continuously upon the parameters a and $\phi(1)$ in the region where $\phi(x)$ is bounded on the interval $1\geq x\geq 0$, it follows that any singularity in the solution of the integral equation must appear first at $x=0$. The nature of the singularity at $x=0$ is easily shown to be

$$\phi(x)\sim\left(\frac{18}{a}\right)^2\frac{1}{x^3}. \tag{34}$$

The solution with this behavior at $x=0$ should then yield the boundary $\phi(1)=f(a)$ of the admissible region. Now defining $\psi(x)$ by

$$\phi(x)=\left(\frac{18}{a}\right)^2\left(\frac{1}{x^3}\right)\psi(x), \tag{35}$$

and substituting this equation into Eq. (32), $\psi(x)$ is seen to satisfy the integral equation

$$\psi(x)=\left(\frac{a}{18}\right)^2\phi(1)x^4+\left(\frac{a}{18}\right)^2 ax^3\int_x^1 dy(y-x)\times yF\left((18/a)^2\frac{\psi(y)}{y^4}\right). \tag{36}$$

This equation has been solved numerically for $\phi(1)\leqq 0$ to obtain the function $\phi(1)=f(a)$, which defines the allowed region. The required solution $\psi(x)$ must be bounded on the interval $0\leq x\leq 1$ and, by virtue of Eqs. (34) and (35), must have the property

$$\psi(0)=1. \tag{37}$$

Figure 1 shows the resultant curve connecting a and $\phi(1)$, which provides the important boundary for delimiting admissible solutions of Eq. (32). The extension to positive values of $\phi(1)$ was obtained by extrapolating the numerical results for $\phi(1)\leq 0$ so as to agree asymptotically with approximate results from Eq. (82). From the latter equation an approximate expression is obtained for the function $\phi(1)=f(a)$ by allowing $\phi(0)$ and, hence, also α (Eq. (72)) to approach infinity. This leads to an explicit relationship between a and $\phi(1)$, where it is necessary in Eq. (82) to make use of the result[10]

$$\psi_0(\alpha)\sim\frac{290}{\alpha^3},$$

which holds approximately for large α.

[10] This result was estimated numerically by examining the behavior of the solutions of the Thomas-Fermi equation for zero temperature.

The physical significance of the boundary curve $\phi(1)=f(a)$ may be seen from the definition of the quantities $\phi(0)$, $\phi(1)$ and a. Since $\phi(0)$ is infinite along this curve, it may be seen from Eq. (6) that either kT or r_0 is zero. As will be shown below a solution $\phi(x)$ is a monotonic increasing function of $\phi(1)$ for fixed x and a, and therefore the admissible region for bounded solutions is defined by $\phi(1)<f(a)$. But by starting at a given point on $\phi(1)=f(a)$ the work function $\phi(1)$ can be made to decrease only by increasing kT or decreasing r_0. Hence, the boundary curve is defined by $kT=0$ and, from the definition of a, by $r_0=\infty$. Physically, the state of the atom is not uniquely defined by the conditions zero temperature and infinite volume. The atomic state under these conditions depends upon the relative rates at which temperature approaches zero and volume approaches infinity. In the present case, this is prescribed by

$$(kT)^{\frac{1}{2}}V^{\frac{1}{3}}\sim\text{constant}.$$

Under this condition, the atom may be shown to approach its ground state as kT approaches zero.

With the assumption that the bounded solutions of Eq. (32) depend sufficiently smoothly on the parameters a and $\phi(1)$ it is possible to demonstrate certain monotonicity properties of the solutions relative to these parameters. These properties are: (i) $\phi(x)$ increases monotonely as a increases provided x and $\phi(1)$ are held fixed, and (ii) $\phi(x)$ increases monotonely as $\phi(1)$ increases provided x and a are held fixed. To prove (i) the derivative of $\phi(x)$ with respect to a can be found from Eq. (32),

$$\frac{\partial\phi(x)}{\partial a}=\int_x^1 dy(y-x)yI_{\frac{1}{2}}\left(\frac{\phi(y)}{y}\right)+a\int_x^1 dy(y-x)I_{\frac{1}{2}}'\left(\frac{\phi(y)}{y}\right)\frac{\partial\phi(y)}{\partial a}. \tag{38}$$

Differentiating with respect to x,

$$\frac{\partial}{\partial x}\frac{\partial\phi(x)}{\partial a}=-\int_x^1 dyyI_{\frac{1}{2}}\left(\frac{\phi(y)}{y}\right)-a\int_x^1 dyI_{\frac{1}{2}}'\left(\frac{\phi(y)}{y}\right)\frac{\partial\phi(y)}{\partial a}, \tag{39}$$

$$\frac{\partial^2}{\partial x^2}\frac{\partial\phi(x)}{\partial a}=xI_{\frac{1}{2}}\left(\frac{\phi(x)}{x}\right)+aI_{\frac{1}{2}}'\left(\frac{\phi(x)}{x}\right)\frac{\partial\phi(x)}{\partial a}. \tag{40}$$

Since

$$\frac{\partial\phi(x)}{\partial a}=\frac{\partial\phi(x)}{\partial a}\bigg|_{x=1}+\frac{\partial}{\partial x}\frac{\partial\phi(x)}{\partial a}\bigg|_{x=1}(x-1)+\frac{\partial^2}{\partial x^2}\frac{\partial\phi(x)}{\partial a}\bigg|_{x=1}\frac{(x-1)^2}{2}+\cdots, \tag{41}$$

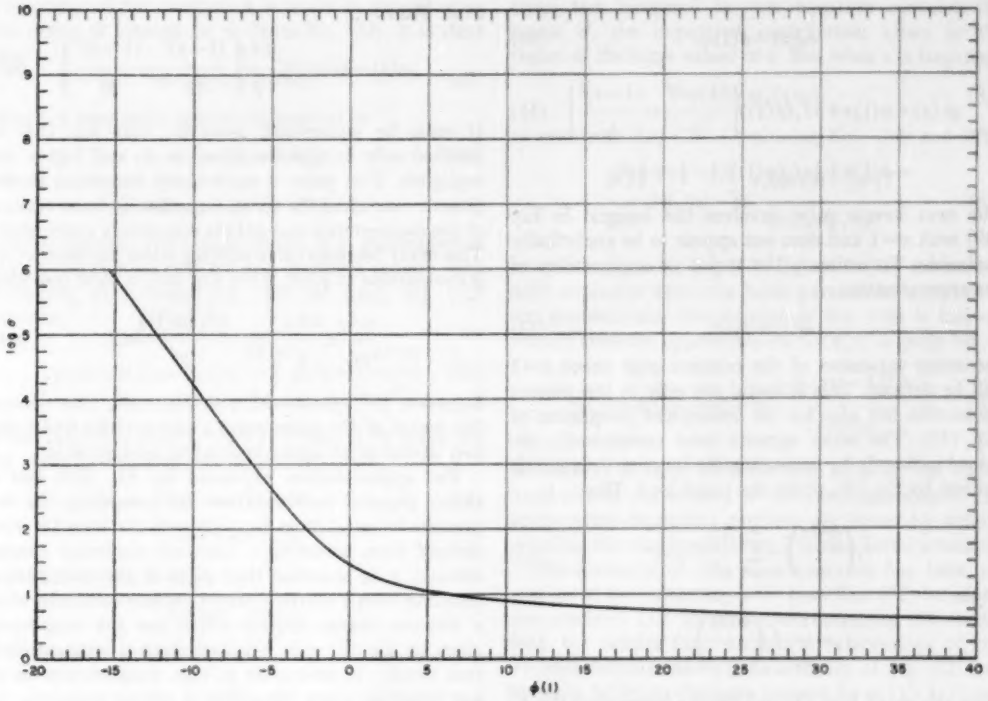


FIG. 1. Boundary curve for the region of admissible solutions of Eq. (32).

it is clear from Eqs. (38), (39), and (40) that

$$\frac{\partial\phi(x)}{\partial a}\sim I_{\frac{1}{2}}(\phi(1))\frac{(x-1)^2}{2}+\cdots. \tag{42}$$

Hence, in a small neighborhood around $x=1$,

$$\frac{\partial\phi(x)}{\partial a}>0. \tag{43}$$

From Eq. (38) it follows that

$$\frac{\partial\phi(x)}{\partial a}>0 \tag{44}$$

everywhere on the interval $0\leq x\leq 1$, since otherwise there would exist a point x_0, $0\leq x_0\leq 1$, for which

$$\frac{\partial\phi(x_0)}{\partial a}=0=\int_{x_0}^{1}dy(y-x_0)yI_{\frac{1}{2}}\left(\frac{\phi(y)}{y}\right)+a\int_{x_0}^{1}dy(y-x_0)I_{\frac{1}{2}}'\left(\frac{\phi(y)}{y}\right)\frac{\partial\phi(y)}{\partial a}. \tag{45}$$

By assumption, however, both integrands in this equation are non-negative on the interval $(x_0,1)$, which denies the existence of the point x_0 and implies that Eq. (44) is true for all x on (0,1). This establishes case (i).

For case (ii), differentiation of Eq. (32) with respect to $\phi(1)$ gives

$$\frac{\partial\phi(x)}{\partial\phi(1)}=x+a\int_{x}^{1}dy(y-x)I_{\frac{1}{2}}'\left(\frac{\phi(y)}{y}\right)\frac{\partial\phi(y)}{\partial\phi(1)}. \tag{46}$$

Since

$$\left.\frac{\partial\phi(x)}{\partial\phi(1)}\right|_{x=1}=1, \tag{47}$$

there exists a neighborhood about $x=1$ for which

$$\partial\phi(x)/\partial\phi(1)>0. \tag{48}$$

The proof that Eq. (48) holds everywhere on the interval $0\leq x\leq 1$ is essentially the same as that given for case (i) and will be omitted.

IV. APPROXIMATE SOLUTIONS OF THE THOMAS-FERMI EQUATION

In this section two approximate solutions for $\phi(x)$ will be developed. They are appropriate to different regions of the $a-\phi(1)$ plane. The first of these concerns the iterates of Eq. (32), given by

$$\phi_{n+1}(x)=\phi(1)x+a\int_{x}^{1}dy(y-x)yI_{\frac{1}{2}}\left(\frac{\phi_n(y)}{y}\right). \tag{49}$$

If

$$\phi_0(x)=\phi(1)x, \tag{50}$$

then

$$\phi_1(x)=\phi(1)x+aI_{\frac{1}{2}}(\phi(1))\left[\frac{(1-x)^2}{2}-\frac{(1-x)^3}{6}\right] \tag{51}$$

$$=\phi(1)x+\tfrac{1}{2}aI_{\frac{1}{2}}(\phi(1))(1-\tfrac{3}{2}x+\tfrac{1}{2}x^2).$$

The next iterate $\phi_2(x)$ involves the integral in Eq. (49) with $n=1$ and does not appear to be analytically evaluable. To estimate the region of applicability of the approximation

$$\phi(x)\approx\phi_1(x), \tag{52}$$

the series expansion of the solution $\phi(x)$ about $x=1$ will be derived. This is useful not only in the present connection but also for the subsequent integration of Eq. (32). The series appears most conveniently obtained indirectly by expanding the function $I_{\frac{1}{2}}(\phi(x)/x)$, defined by Eq. (4), about the point $x=1$. Thus

$$I_{\frac{1}{2}}\left(\frac{\phi(x)}{x}\right)=\sum_{n=0}^{\infty}\frac{A_n}{n!}(1-x)^n, \tag{53}$$

where

$$A_n=\frac{d^n}{dx^n}I_{\frac{1}{2}}\left(\frac{\phi(x)}{x}\right)\bigg|_{x=1}. \tag{54}$$

With the aid of Eqs. (3) and (5) the expressions for A_n can be evaluated. In particular

$$A_0=I_{\frac{1}{2}}(\phi(1)), \tag{55}$$

$$A_1=0, \tag{56}$$

$$A_2=aI_{\frac{1}{2}}'(\phi(1))\,I_{\frac{1}{2}}(\phi(1)), \tag{57}$$

$$A_3=2aI_{\frac{1}{2}}'(\phi(1))\,I_{\frac{1}{2}}(\phi(1)), \tag{58}$$

$$A_4=3a^2I_{\frac{1}{2}}''(\phi(1))\,I_{\frac{1}{2}}^2(\phi(1))+a^2I_{\frac{1}{2}}'^2(\phi(1))\,I_{\frac{1}{2}}(\phi(1)) +8aI_{\frac{1}{2}}'(\phi(1))\,I_{\frac{1}{2}}(\phi(1)), \tag{59}$$

$$A_5=20a^2I_{\frac{1}{2}}''(\phi(1))\,I_{\frac{1}{2}}^2(\phi(1))+4a^2I_{\frac{1}{2}}'^2(\phi(1))\,I_{\frac{1}{2}}(\phi(1)) +40aI_{\frac{1}{2}}'(\phi(1))\,I_{\frac{1}{2}}(\phi(1)), \tag{60}$$

$$A_6=15a^3I_{\frac{1}{2}}'''(\phi(1))\,I_{\frac{1}{2}}^3(\phi(1))+200a^2I_{\frac{1}{2}}''(\phi(1))\,I_{\frac{1}{2}}^2(\phi(1)) +18a^3I_{\frac{1}{2}}''(\phi(1))\,I_{\frac{1}{2}}'(\phi(1))\,I_{\frac{1}{2}}^2(\phi(1)) +24a^2I_{\frac{1}{2}}'^2(\phi(1))\,I_{\frac{1}{2}}(\phi(1))+a^3I_{\frac{1}{2}}'^3(\phi(1))\,I_{\frac{1}{2}}(\phi(1)) +240a^2I_{\frac{1}{2}}'(\phi(1))\,I_{\frac{1}{2}}(\phi(1)). \tag{61}$$

Substitution of Eq. (53) into Eq. (32) leads to

$$\phi(x)=\phi(1)x +a\sum_{k=0}^{\infty}\frac{A_k}{k!}\left[\frac{(1-x)^{k+2}}{(k+1)(k+2)}-\frac{(1-x)^{k+3}}{(k+2)(k+3)}\right]. \tag{62}$$

Comparison of this result with Eq. (51) shows that the approximation $\phi_1(x)$ is equivalent to approximating the series by the A_0 term only. Since $A_1=0$, the term A_2 measures the accuracy with which $\phi_1(x)$ represents $\phi(x)$. Thus

$$|\phi(x)-\phi_1(x)|\leq\frac{aA_2}{2}\left[\frac{(1-x)^4}{12}-\frac{(1-x)^5}{20}\right]. \tag{63}$$

It must be understood, however, that Eq. (63) is justified only in case the terms in A_2 and higher are negligible. This point is particularly important in the present case since the series expansion in some regions of the parameters a and $\phi(1)$ is very slowly convergent. This effect becomes quite striking when the term in A_2 is comparable to $\phi_1(x)$. Thus Eq. (63) is valid provided

$$\frac{aA_2}{60}\ll\frac{aA_0}{3} \quad\text{or}\quad \frac{aI_{\frac{1}{2}}'(\phi(1))}{20}\ll 1. \tag{64}$$

Equation (57), combined with Eq. (64), thus defines the region of the parameters a and $\phi(1)$ for which the first iterate $\phi_1(x)$ approximates the solution $\phi(x)$.

The approximation expressed by Eq. (52) has a simple physical interpretation. By comparing the expression for $\phi_1(x)$ from Eq. (51) with the form for $\phi(x)$ derived from assuming a constant electronic charge density, it is observed that $\phi_1(x)$ is the distribution resulting from a central charged nucleus combined with a uniform charge density which has the magnitude given by Eq. (7) with $V(r_0)=0$. Such a constant electron density is reasonable at high temperatures or at low densities where ionization is almost complete. In this limit the thermodynamic functions are easily evaluated from Eqs. (10), (12), (15), and (31). Equations (10) and (31) are unchanged, but Eqs. (12) and (15) simplify to

$$E_{\text{kin}}=\left(\frac{ZkT}{3}\right)\left[\frac{a}{\phi(0)}\right]I_{\frac{1}{2}}(\phi(1)), \tag{12'}$$

and

$$E_{\text{pot}}=-ZkT\left(\frac{11}{20}\right)aI_{\frac{1}{2}}(\phi(1)), \tag{15'}$$

where $\phi(0)=\frac{1}{3}aI_{\frac{1}{2}}(\phi(1))$, as given by Eqs. (51) and (52).

While this approximation gives $\phi(x)$ with reasonable accuracy over a wide range of the parameters a and $\phi(1)$, the energies as specified by Eqs. (12′) and (15′) may be inaccurate. A comparison of this approximation with the numerical results obtained below showed that even if $\phi(1)$ were accurate to one percent, Eq. (12′) was inaccurate by as much as a factor of 1.3 and Eq. (15′) by a factor of 30. However, E_{pot} may be obtained from the virial theorem using the more accurate E_{kin} from Eq. (12′).

Another useful approximation can be obtained if the restriction is made that

$$\phi(1)\geq 0. \tag{65}$$

In this case the function $I_{\frac{1}{2}}(\eta)$ of Eq. (4) is approximated by

$$I_{\frac{1}{2}}(\eta)\sim A+\tfrac{2}{3}\eta^{\frac{3}{2}}. \tag{66}$$

The choice of the constant A is made to depend upon the region of interest for η. From Eq. (32), it is clear that

$$\eta=\phi(y)/y\geq\phi(1). \tag{67}$$

Hence, a reasonable choice[11] appears to be

$$A=I_{\frac{1}{2}}(\phi(1))-\tfrac{2}{3}\phi^{\frac{3}{2}}(1); \tag{68}$$

that is,

$$I_{\frac{1}{2}}(\eta)\sim\{I_{\frac{1}{2}}(\phi(1))-\tfrac{2}{3}\phi^{\frac{3}{2}}(1)\}+\tfrac{2}{3}\eta^{\frac{3}{2}}. \tag{69}$$

The accuracy of this approximation is in general better than 5 percent for $\phi(1)>0$ and increases rapidly with increasing $\phi(1)$. Using Eq. (69) for $I_{\frac{1}{2}}(\eta)$, Eq. (32) becomes

$$\phi(x)=\phi(1)x+\gamma(x)+\frac{2a}{3}\int_x^1 dy(y-x)\frac{\phi^{\frac{3}{2}}(y)}{y^{\frac{1}{2}}} \tag{70}$$

where $\gamma(x)=(aA/3)(1-\frac{3}{2}x+\frac{1}{2}x^3)$ and A is given by Eq. (68). In terms of the quantities

$$\psi(ax)=\phi(x)/\phi(0), \tag{71}$$

$$\alpha=[\tfrac{2}{3}a\phi^{\frac{1}{2}}(0)]^{\frac{2}{3}}, \tag{72}$$

and by changing scale to

$$u=\alpha x, \quad v=\alpha y, \tag{73}$$

Eq. (70) may be expressed as

$$\psi(u)=\psi(\alpha)\frac{u}{\alpha}+\frac{\gamma(u/\alpha)}{\phi(0)}+\int_u^\alpha dv(v-u)\frac{\psi^{\frac{3}{2}}(v)}{v^{\frac{1}{2}}}. \tag{74}$$

Except for the second term on the right-hand side, this equation is identical with the Thomas-Fermi equation for zero temperature, which suggests that a perturbation solution be looked for. Let

$$\psi(u)=\psi_0(u)+\epsilon(u), \tag{75}$$

where

$$\psi_0(u)=\psi_0(\alpha)\frac{u}{\alpha}+\int_u^\alpha dv(v-u)\frac{\psi_0^{\frac{3}{2}}(v)}{v^{\frac{1}{2}}}. \tag{76}$$

If it is assumed that

$$\psi_0(0)=\psi(0)=1, \tag{77}$$

then

$$\epsilon(0)=0, \quad \epsilon'(\alpha)=\epsilon(\alpha)/\alpha, \tag{78}$$

and

$$\epsilon(u)=\epsilon(\alpha)\frac{u}{\alpha}+\frac{\gamma(u/\alpha)}{\phi(0)}+\int_u^\alpha dv(v-u)\frac{\psi^{\frac{3}{2}}(v)-\psi_0^{\frac{3}{2}}(v)}{v^{\frac{1}{2}}}. \tag{79}$$

An approximate solution to this equation is obtained by iteration, assuming, to begin with, that $\epsilon(v)$ in the integrand may be replaced by $\epsilon(\alpha)(v/\alpha)$. Moreover, if $\epsilon(u)\ll\psi_0(u)$,

$$0=\frac{\gamma(0)}{\phi(0)}+\frac{3}{2}\frac{\epsilon(\alpha)}{\alpha}\int_0^\alpha dv v^{\frac{1}{2}}\psi_0^{\frac{1}{2}}(v). \tag{80}$$

[11] This form was suggested by D. Liberman.

Since the integrand in this equation contains the factor $v^{\frac{1}{2}}$, the important contribution arises in the region of the large values of v. But when v is large,

$$\psi_0(v)\sim\psi_0(\alpha)v/\alpha \tag{81}$$

as seen from Eq. (76). Combining Eqs. (80) and (81),

$$\frac{\phi(1)}{\phi(0)}=\psi_0(\alpha)-\tfrac{2}{3}a\frac{I_{\frac{1}{2}}(\phi(1))-\frac{2}{3}\phi^{\frac{3}{2}}(1)}{\phi(0)\alpha^{\frac{1}{2}}\psi_0^{\frac{1}{2}}(\alpha)}, \tag{82}$$

where α is given by Eq. (72). Equation (82) determines $\phi(0)$ implicitly when the basic parameters a and $\phi(1)$ are specified and the solution of Eq. (76) is known. Finally then the approximation for $\phi(x)$ is given by

$$\phi(x)\approx\phi(0)\psi_0(\alpha x)-\tfrac{2}{3}a\frac{I_{\frac{1}{2}}(\phi(1))-\frac{2}{3}\phi^{\frac{3}{2}}(1)}{\alpha^{\frac{1}{2}}\psi_0^{\frac{1}{2}}(\alpha)}x. \tag{83}$$

When $\phi(1)$ is large, only the first term on the right-hand side is important; in this case, solutions of the temperature-dependent problem are found by simply rescaling the zero temperature Thomas-Fermi solutions.

The accuracy of this approximation has been investigated by computing $\phi(1)$ from Eq. (82) for given temperatures and volumes and comparing the results with the results from numerical integrations of the Thomas-Fermi equation. The accuracy of Eq. (82) was found to be better than one percent for $\phi(1)>4$; it was about two percent for $\phi(1)>2$ and better than ten percent for $\phi(1)>1$. For temperatures and volumes corresponding to $\phi(1)=0$, Eq. (82) predicted that $|\phi(1)|<0.3$.

The pressure is calculated in this approximation by a straightforward use of Eqs. (82) and (10). On the other hand, the energies and the entropy require integrations involving the solutions $\psi_0(\alpha x)$ of the zero-temperature Thomas-Fermi equation, which integrations have not been carried out in view of the non-approximate numerical calculations discussed below.

V. NUMERICAL SOLUTION OF THE THOMAS-FERMI EQUATION

The numerical scheme for treating Eq. (32) consists in replacing the integral in the latter equation by a k-point numerical integration formula. This was taken to be the ordinary Simpson's rule, which is a three-point formula and has the particular advantage that the numerical accuracy is in the fourth power of the interval size.

Before describing further the details of the numerical formulation, it is necessary to make some remarks about the nature of the functions $\phi(x)$ and $I_{\frac{1}{2}}(\eta)$. In the first place, the function $\phi(x)$, which is directly related to the potential distribution within the atom, changes most rapidly in the neighborhood of the origin since, in general, the greatest charge density occurs there. It seemed desirable, therefore, to emphasize the

neighborhood of $x=0$ by making a change of variable, namely,

$$x=u^2 \quad \text{and} \quad y=v^2. \tag{84}$$

In these variables, Eq. (32) becomes

$$\phi(u)=\phi(1)u^2+2a\int_u^1 dv(v^2-u^2)v^2 I_{\frac{1}{2}}\left(\frac{\phi(v)}{v^3}\right). \tag{85}$$

The division points for the numerical integration are then made on the variable u.

The next point concerns the function $I_{\frac{1}{2}}(\eta)$, which is defined in Eq. (4). Tables of this function for a large range of its argument were obtained by Stoner and McDougall.[12] By numerical fitting of the tables with polynomials and by using asymptotic expansions for $I_{\frac{1}{2}}(\eta)$ outside the range of the tables, a set of simple analytic forms which are conveniently handled numerically were derived. Explicitly, these forms are:

(i) $-\infty<\eta\leq -2.0$,

$$I_{\frac{1}{2}}(\eta)=\frac{\sqrt{\pi}}{2}e^{\eta}\left[1-\frac{e^{\eta}}{2^{\frac{1}{2}}}+\frac{e^{2\eta}}{3^{\frac{1}{2}}}-\frac{e^{3\eta}}{4^{\frac{1}{2}}}+\frac{e^{4\eta}}{5^{\frac{1}{2}}}-\frac{e^{5\eta}}{6^{\frac{1}{2}}}+\frac{e^{6\eta}}{7^{\frac{1}{2}}}\right], \tag{86}$$

(ii) $-2.0\leq\eta\leq 0$,

$$I_{\frac{1}{2}}(\eta)=0.678091+0.53619667\eta+0.16909748\eta^2+0.018780823\eta^3-0.002357546\eta^4-0.000639610797\eta^5, \tag{87}$$

(iii) $0\leq\eta\leq 3.0$,

$$I_{\frac{1}{2}}(\eta)=0.678091+0.53638\eta+0.1668235\eta^2+0.0206067\eta^3-0.00601491\eta^4+0.000490398\eta^5, \tag{88}$$

(iv) $3.0\leq\eta\leq 10.0$,

$$I_{\frac{1}{2}}(\eta)=0.757064709+0.3922888\eta+0.2705525\eta^2-0.01682933\eta^3+0.0008258364\eta^4-0.000001819771\eta^5, \tag{89}$$

(v) $10.0\leq\eta\leq 10^5$,

$$I_{\frac{1}{2}}(\eta)=\tfrac{2}{3}\eta^{\frac{3}{2}}\left[1+\frac{1.2337005}{\eta^2}+\frac{1.0654119}{\eta^4}+\frac{9.7015185}{\eta^6}+\frac{242.71502}{\eta^8}+\frac{12313.691}{\eta^{10}}\right], \tag{90}$$

(vi) $10^5\leq\eta<\infty$,

$$I_{\frac{1}{2}}(\eta)=\tfrac{2}{3}\eta^{\frac{3}{2}}. \tag{91}$$

The fractional error in the fitting of these functions to Eq. (4) was estimated from the tables given by Stoner and McDougall to be at most 6×10^{-5} and, except for a few isolated regions, less than 1×10^{-5}.

Finally, it is necessary to stipulate the accuracy with which $\phi(x)$ must be calculated. Since the Thomas-Fermi model of an atomic system is only a crude approximation to an actual system, it might seem at first that rough accuracy for $\phi(x)$ would be sufficient. A difficulty arises, however, in determining the total energy of the system. An atomic system has, in its ground state, total energies varying from about 10 ev to about 10^6 ev. In estimating the effect of temperature and pressure, it is necessary to evaluate the change in the total energy of the atom in proceeding from one state to a neighboring one. The important physical quantity is this change in the energy, which in cases of interest may be as small as 1 to 10 ev. This implies that for atoms with the largest total binding energies, accuracies of the order of one part in 10^7 to 10^8 are necessary. Moreover, additional figures must be carried in the individual component numbers in order to absorb the inevitable rounding and truncation errors. In the present work, ten figures were employed on all numbers. To insure the desired accuracy of one part in 10^8, the numerical integration of Eq. (85) was repeated, halving each time the interval size in the integration scheme. It is clear, of course, that two such successive integrations need only agree to 2 parts in 10^7 since in using Simpson's rule the error decreases with the fourth power of the interval size; if two integrations agree to 2×10^{-7}, the one with the smaller interval size will have the fractional accuracy of 10^{-8}.

In accordance with the previous remarks, $\phi(x)$ is to be determined from the difference equations corresponding to Eq. (85), replacing the integral by a Simpson's rule sum. In particular if the division points for the integration are denoted by u_n and if the subscript n is used to indicate that a function is evaluated at u_n, then the basic difference equations become

$$\phi_n=\phi(1)u_n^2+2a[Q_n-u_n^2P_n], \tag{92}$$

where

$$Q_n=\frac{\Delta u}{3}[g_n+4g_{n+1}+g_{n+2}]+Q_{n+2}, \tag{93}$$

$$P_n=\frac{\Delta u}{3}[f_n+4f_{n+1}+f_{n+2}]+P_{n+2}, \tag{94}$$

$$g_n=u_n^5 I_{\frac{1}{2}}(\phi_n/u_n^3), \tag{95}$$

$$f_n=u_n^3 I_{\frac{1}{2}}(\phi_n/u_n^3), \tag{96}$$

or

$$\phi_n=\phi(1)u_n^2+2a\{\tfrac{1}{3}\Delta u[(4g_{n+1}+g_{n+2})-u_n^2(4f_{n+1}+f_{n+2})]+Q_{n+2}-u_n^2P_{n+2}\}, \tag{97}$$

where $\Delta u=1/N$ and N is the total numbers of integration points. To initiate the solution of Eq. (97), it is necessary to specify $\phi(u)$, $P(u)$ and $Q(u)$ at the first two points of the integration. These can be evaluated from Eq. (62). Namely,

(i) $\phi_N=\phi(1)$,
$Q_N=0$,
$P_N=0$,

[12] J. McDougall and E. C. Stoner, Phil. Trans. Roy. Soc. (London) **237**, 67 (1938).

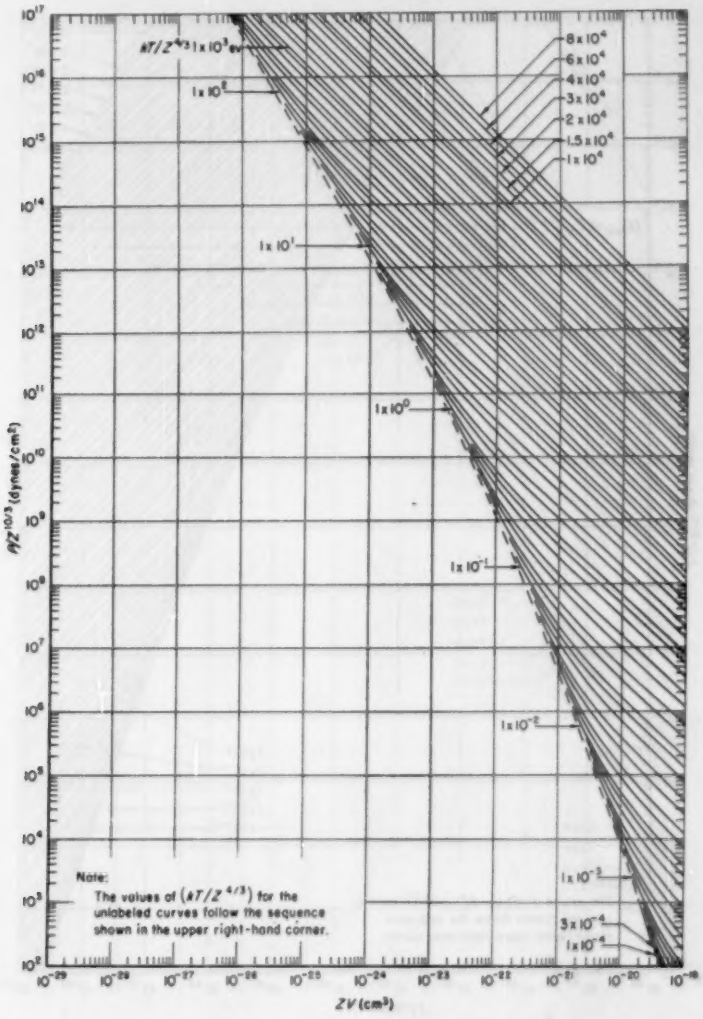


FIG. 2. Pressure *versus* volume for fixed temperature.

(ii) $$\phi_{N-1}=\phi(1)u_{N-1}{}^2+a\sum_{k=0}^{2}\frac{A_k}{k!}\times\left[\frac{(1-u_{N-1}{}^2)^{k+2}}{(k+1)(k+2)}-\frac{(1-u_{N-1}{}^2)^{k+3}}{(k+2)(k+3)}\right],$$

$$P_{N-1}=\tfrac{1}{2}\sum_{k=0}^{2}\frac{A_k}{k!}\left[\frac{(1-u_{N-1}{}^2)^{k+1}}{k+1}-\frac{(1-u_{N-1}{}^2)^{k+2}}{k+2}\right],$$

$$Q_{N-1}=-\tfrac{1}{2}\sum_{k=0}^{2}\frac{A_k}{k!}\left[\frac{(1-u_{N-1}{}^2)^{k+2}}{k+2}-\frac{(1-u_{N-1}{}^2)^{k+3}}{k+3}\right]+\tfrac{1}{2}\sum_{k=0}^{2}\frac{A_k}{k!}\left[\frac{(1-u_{N-1}{}^2)^{k+1}}{k+1}-\frac{(1-u_{N-1}{}^2)^{k+2}}{k+2}\right],$$

where the coefficients A_k are defined in Eqs. (55)–(58). Terms up to A_2 only are needed to give accuracy consistent with Simpson's rule.

The integration procedure itself is immediately clear from the difference equations. However, it was found convenient to permit changes of interval size in the course of the computation, which necessitates interpolation for intermediate values of $\phi(u)$, $P(u)$, and $Q(u)$. Cubic interpolation had to be used in order that the error be again consistent with the Simpson's formula.

VI. NUMERICAL RESULTS

The calculations, described in the preceding section, were carried out on the IBM 701 Defense Calculator. Solutions to Eq. (85) were obtained for about 1000 sets of values of the parameters a and $\phi(1)$, covering a

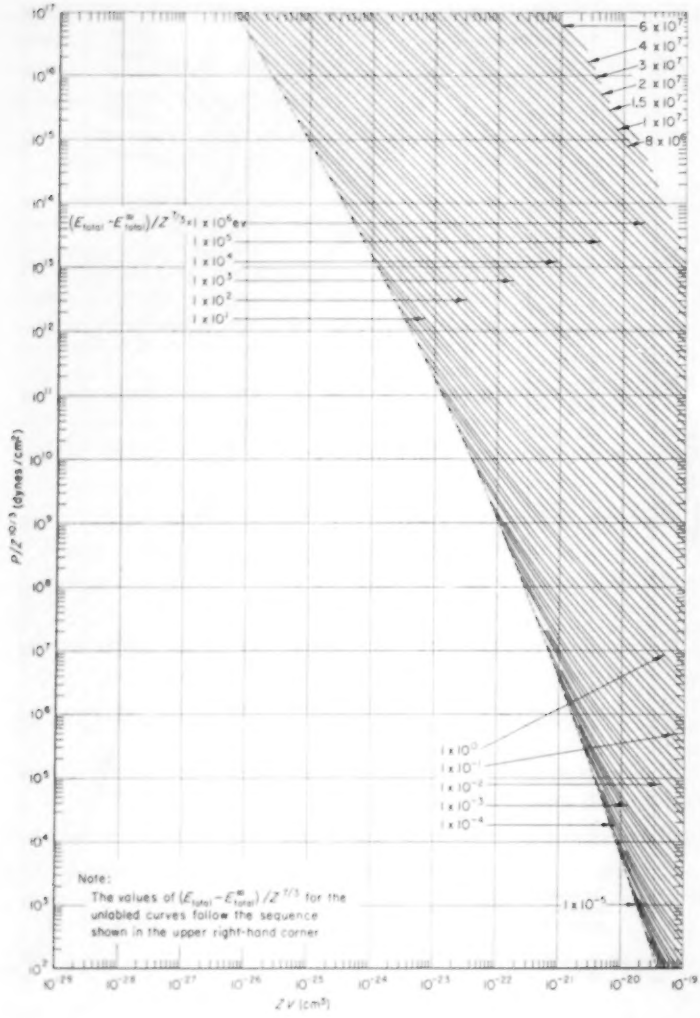


FIG. 3. Pressure *versus* volume for fixed total energy.

range of temperatures from about 10^{-4} ev–10^3 ev. The particular values of a and $\phi(1)$ were selected to encompass a reasonably complete range of physically interesting atomic states and for ease of interpolation. The resultant physical data[13] are presented in Figs. 2 to 6. A more precise presentation of the data does not appear justified in view of the approximate character of the model.

For clarity in the data, the solutions obtained for particular values of a and $\phi(1)$ were interpolated to intermediate solutions with prescribed values of temperature, energy, or entropy. This procedure involved quadratic interpolation on the basic solutions and consequently introduced a significant depreciation in the accuracy of the data. A check was made on the final accuracy of the data and was found to be in general better than one-tenth percent, with some reservation with respect to the energies and entropy at very low temperatures. The entropy which is graphed in Fig. 4 involves, as can be seen from Eq. (31), the calculation of differences between large numbers, which differences approach zero as the temperature approaches zero. Thus because of the limitation on the number of digits which are carried within the machine, the entropy decreases in numerical accuracy with decreasing temperature. This comment also applies to the quantities $\Delta E_{\text{tot}}/Z^{7/3}$ and $\Delta E_{\text{kin}}/Z^{7/3}$ which are graphed in Figs. 3,

[13] The physical parameters employed in these calculations were $m=9.10710\times10^{-28}$ g, $e=4.80217\times10^{-10}$ esu, and $h=6.62363\times10^{-27}$ erg-sec. The conversion factor between electron volts and ergs was taken to be 1.6020×10^{-12}.

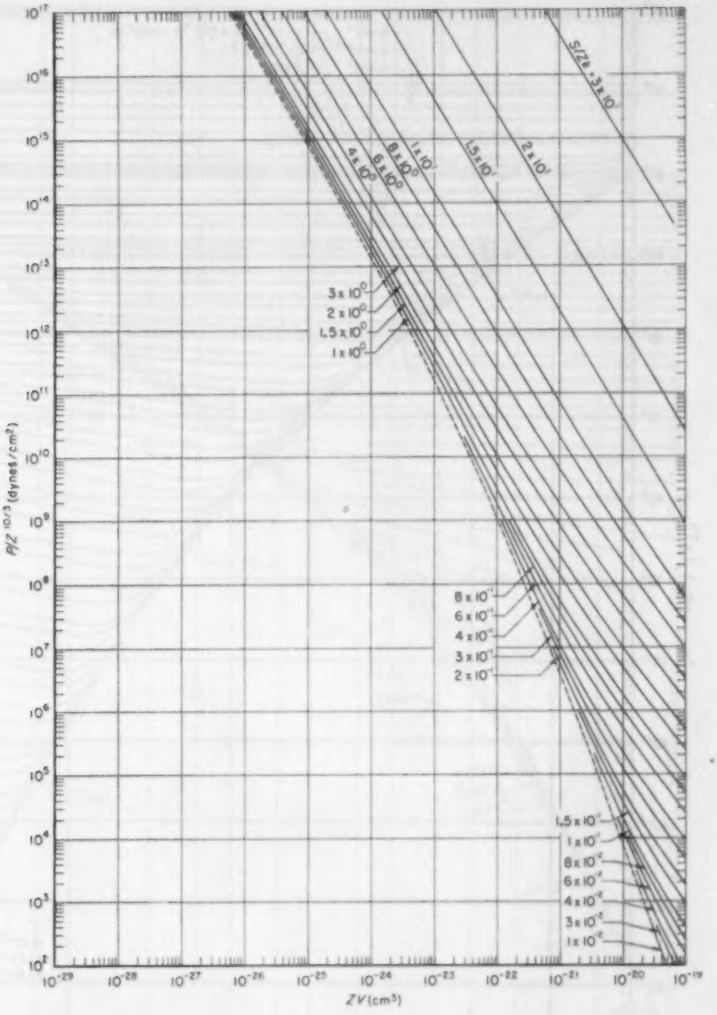


FIG. 4. Pressure *versus* volume for fixed entropy.

5, and 6.[14] These quantities are the differences in the energies of the atom in the prescribed state of temperature and volume and the energies of the atom in its isolated state at zero temperature. $\Delta E_{pot}/Z^{7/3}$ and the ungraphed region of $\Delta E_{kin}/Z^{7/3}$ may be obtained approximately from Figs. 2 and 5 by using the virial theorem of Eq. (16). To complete the definition of these energies, it is observed that

$$E^{x}{}_{tot}/Z^{7/3} = -E^{x}{}_{kin}/Z^{7/3} = \tfrac{1}{2}E^{x}{}_{pot}/Z^{7/3} = -20.913065 \text{ ev}.$$

The evaluation of these numbers was carried out numerically for the isolated atom at zero temperature. The pressure does not involve differences of large numbers and, consequently, has a uniform accuracy. Similarly, the basic parameter $\phi(0)$, which is shown in Fig. 7, is in general accurate to about one-tenth percent.

There exist some regions of the parameters a and $\phi(1)$ which were not explored, either because they were in physically uninteresting regions or because the states in these regions were negligibly different from the zero temperature states. Also the potential distributions $\phi(x)$, which were determined in the calculations, comprise too large a body of numbers to present conveniently. However, with the aid of the parameter values, a and $\phi(1)$, and $\phi(0)$ of Fig. 7, and the integration formulas, a relatively accurate and simple calculation for $\phi(x)$ may be carried out manually in any specific case.

As a final point, a comparison of the present exact

[14] The irregularity of Fig. (6) results from the existence of a region of small negative $\Delta E_{kin}/Z^{7/3}$ which is not conveniently graphed.

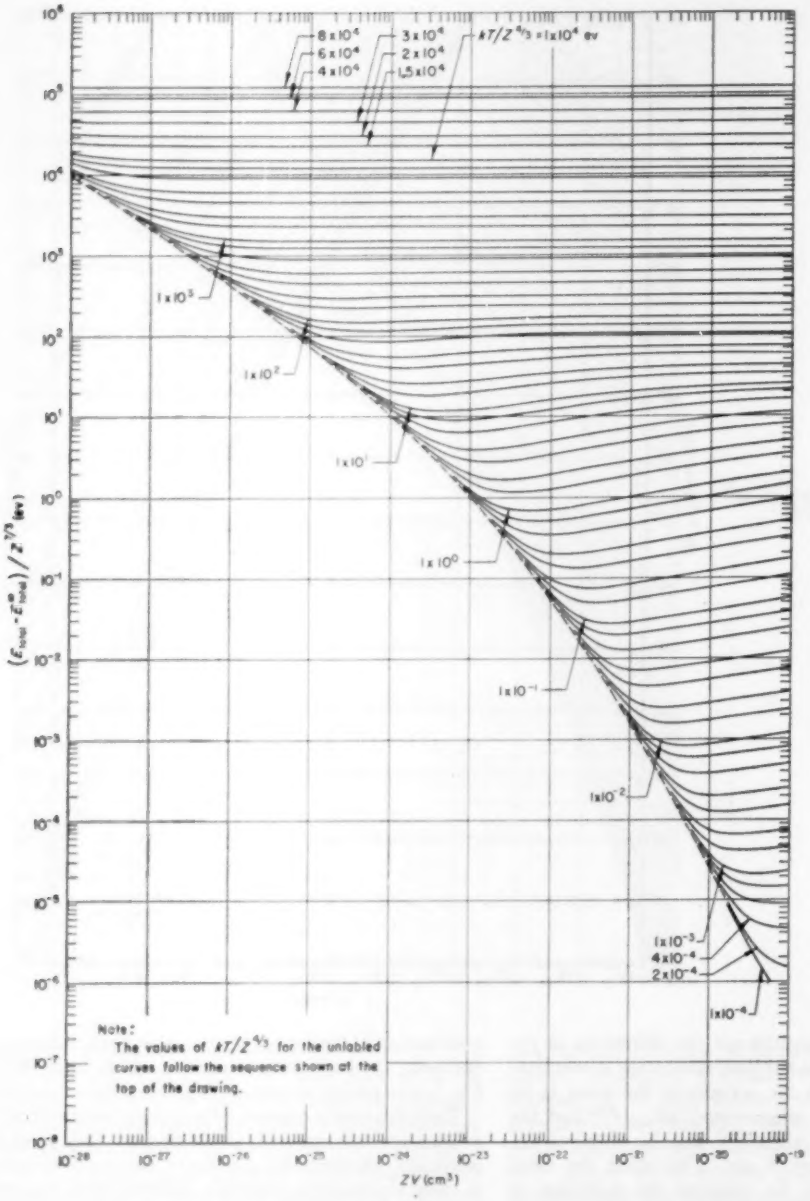


FIG. 5. Total energy *versus* volume for fixed temperature.

results with the results of the temperature-perturbation procedure[3] and with the calculations of Feynman, Metropolis, and Teller[2] has been made. In Figs. 8 and 9, pressures and energies are compared with those derived from the temperature-perturbation procedure, and in Figs. 10 and 11 this same comparison is presented in terms of the scaled variables of that method. The domain of applicability of the perturbation approach is limited to work-functions $\phi(1)$ large compared with unity as a consequence of the asymptotic expansion employed in that method. This region lies below the curves labeled $\phi(1)=1$ in Figs. 8 and 9. It may be observed from Figs. 8 and 9 that the temperature-perturbation method gives satisfactory accuracy in a region which differs by a factor less than about $1\frac{1}{2}$ in pressure and by a factor less than about $1\frac{1}{2}$ to 3 in

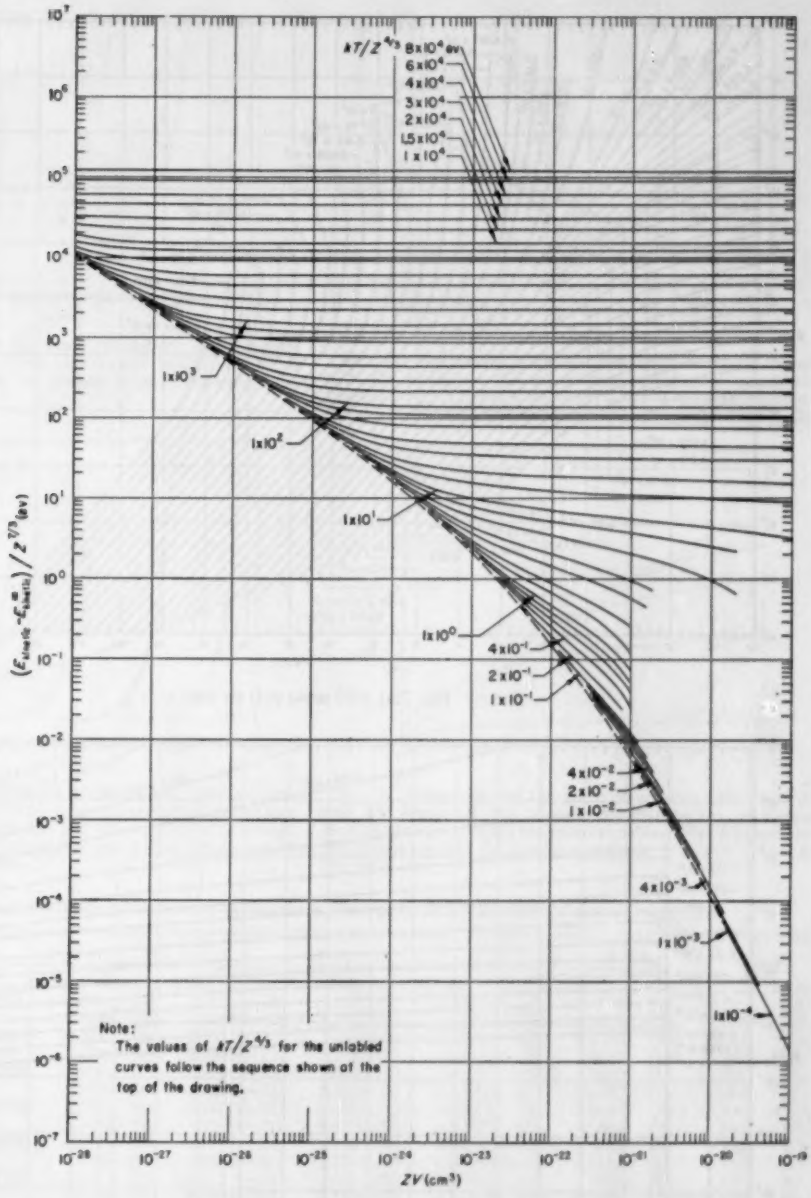


FIG. 6. Kinetic energy *versus* volume for fixed temperature.

energy from the zero-temperature results. When the deviation from the zero-temperature behavior becomes larger, the perturbation results, not unexpectedly, of course, involve sizable errors. Finally, Table I compares the results of Feynman, Metropolis, and Teller (Table XI of reference 2) with the present calculations. This comparison was made by repeating, with the above discussed numerical procedure, the integrations of Feynman, Metropolis, and Teller for the parameter values a and $\phi(1)$ corresponding to their values of $b=a^3$ and $\beta_0=b\phi(1)$. Since the accuracy of the present calculations is one part in 10^7 to 10^8, the discrepancies in the two sets of data appear to be due to typographical errors or small numerical inaccuracies in the integrations by Feynman, Metropolis, and Teller.

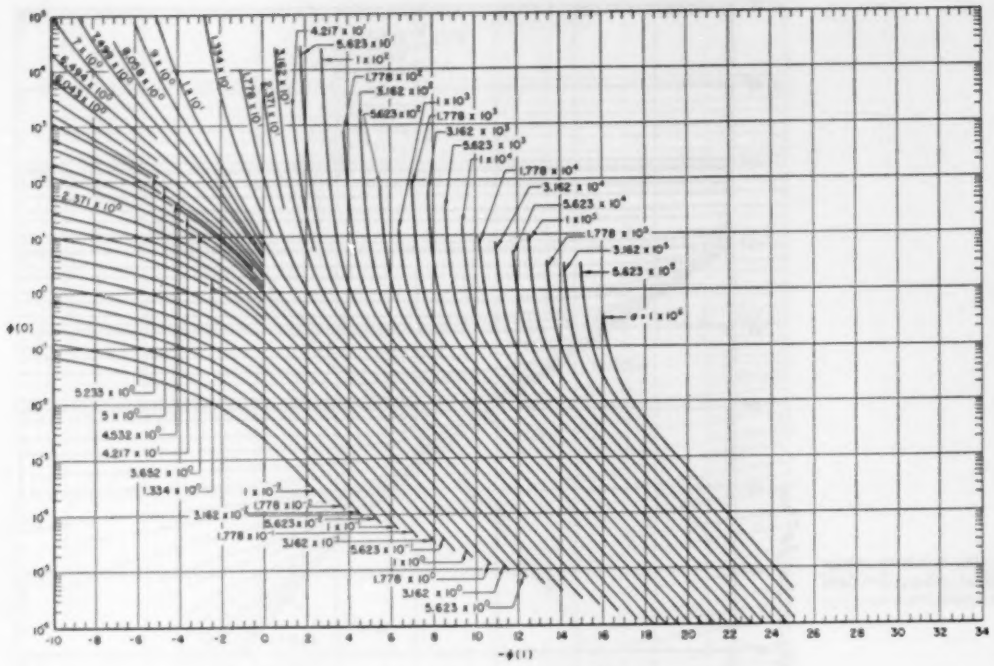


FIG. 7(a). $\phi(0)$ *versus* $\phi(1)$ for fixed a.

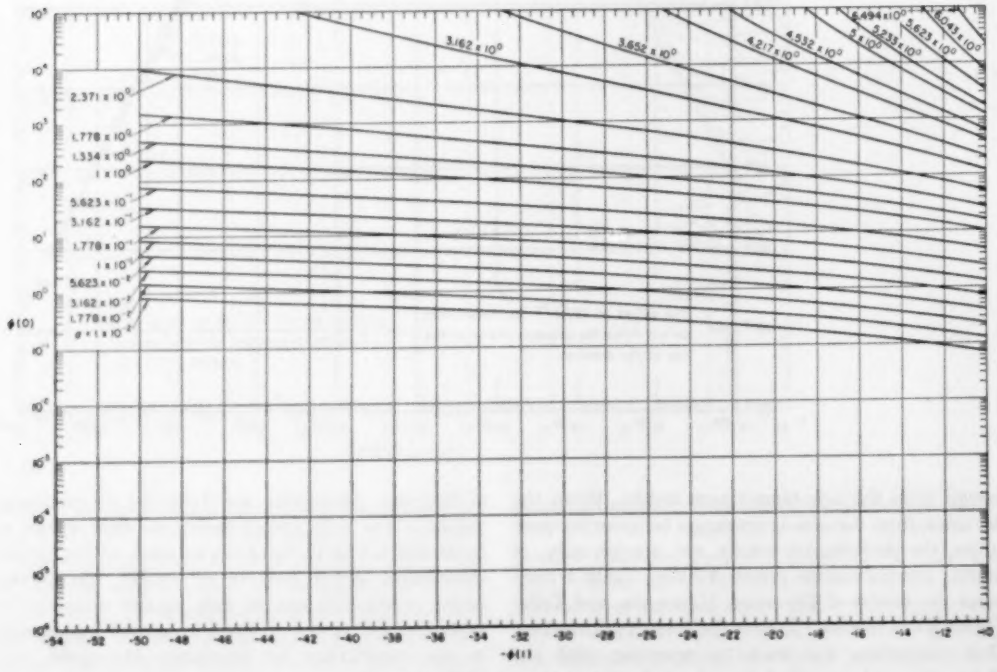


FIG. 7(b). $\phi(0)$ *versus* $\phi(1)$ for fixed a.

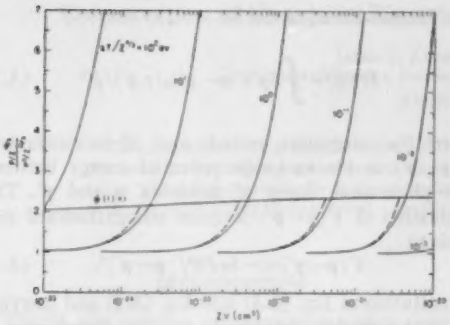


FIG. 8. Comparison of the exact calculations (solid curves) with the results of the temperature-perturbation calculation (dashed curve). P^0 denotes the pressure at zero temperature and at the volume ZV.

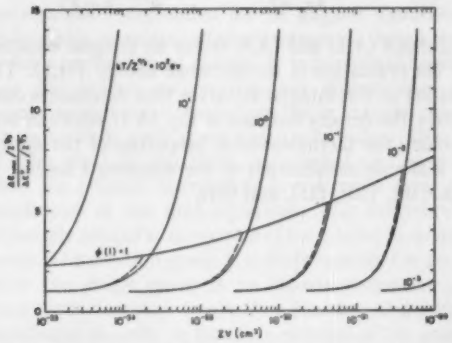


FIG. 9. Comparison of the exact calculations (solid curves) with the results of the temperature-perturbation calculation (dashed curve). $\Delta E_{\text{total}}{}^0$ denotes the difference in total energy between the state of zero temperature and infinite volume and of zero temperature and volume ZV.

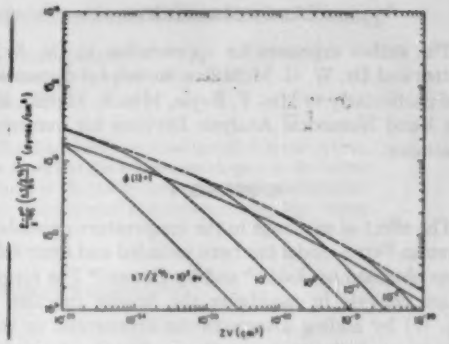


FIG. 10. Comparison of the exact calculations (solid curves) with the results of the temperature-perturbation calculation (dashed curve) in terms of the scaled variables of the temperature-perturbation model. Rh denotes a rydberg; P^0 is defined in Fig. (8).

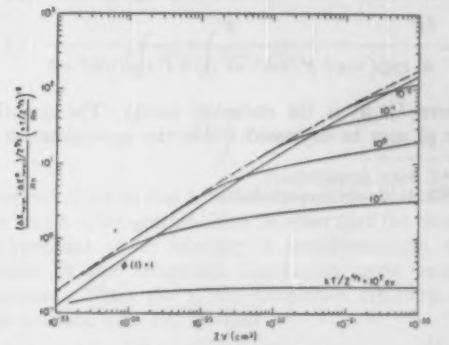


FIG. 11. Comparison of the exact calculations (solid curves) with the results of the temperature-perturbation calculation (dashed curve) in terms of the scaled variables of the temperature-perturbation model. Rh denotes a rydberg; $\Delta E_{\text{total}}{}^0$ is defined in Fig. (9).

TABLE I.[a] Comparison of the calculations of (a) Feynman, Metropolis, and Teller with (b) present calculations.

		a	$\phi(1)$	$\frac{kT}{Z^{4/3}}$(ev)	ZV(cm³)	$\frac{P}{Z^{10/3}}$(dynes/cm²)	$\frac{E_{\text{total}}}{ZkT}$	$\frac{E_{\text{potential}}}{ZkT}$	$\frac{E_{\text{kinetic}}}{ZkT}$
1	(a)	29.1600	−2.5000	2.896 (0)	2.172 (−22)	1.117 (10)	−6.0035 (0)	1.35750 (1)	7.5715 (0)
	(b)			2.8946 (0)	2.1710(−22)	1.1163(10)	−6.0037 (0)	1.3575 (1)	7.5716 (0)
2	(a)	44.3716	−3.43653	6.395 (0)	2.251 (−22)	3.198 (10)	−1.672 (0)	5.4520 (0)	3.780 (0)
	(b)			6.3726 (0)	2.2546(−22)	3.1742(10)	−1.6841 (0)	5.4713 (0)	3.7872 (0)
3	(a)	132.710	−4.48090	1.916 (0)	2.8740(−21)	5.564 (8)	−9.319 (0)	2.02006 (1)	1.0882 (1)
	(b)			1.9056 (0)	2.8840(−21)	5.4816 (8)	−9.3868 (0)	2.0327 (1)	1.0940 (1)
4	(a)	182.790	−5.10000	3.091 (0)	3.2477(−21)	9.872 (8)	−4.859 (0)	1.16576 (1)	6.799 (0)
	(b)			3.0884 (0)	3.2455(−21)	9.8789 (8)	−4.8592 (0)	1.1663 (1)	6.8033 (0)
5	(a)	245.862	−5.90000	6.877 (0)	2.7789(−21)	3.282 (9)	−8.75 (−1)	4.2351 (0)	3.360 (0)
	(b)			6.8672 (0)	2.7804(−21)	3.2745 (9)	−8.7978(−1)	4.2423 (0)	3.3625 (0)
6	(a)	17.6820	−1.77256	3.072 (0)	9.812 (−23)	2.641 (10)	−5.7104 (0)	1.30006 (1)	7.2902 (0)
	(b)			3.0284 (0)	9.9094(−23)	2.5496(10)	−5.8156 (0)	1.3194 (1)	7.3780 (0)
7	(a)	85.4700	−6.63001	1.9033 (2)	4.722 (−23)	3.188 (12)	7.066 (−1)	6.76 (−2)	7.742 (−1)
	(b)			1.9067 (2)	4.7115(−23)	6.4121(12)	1.4528 (0)	6.1584 (−2)	1.5143 (0)
8	(a)	67.2597	−3.79003	3.795 (0)	6.2122(−22)	6.105 (7)	−3.9075 (0)	9.6867 (0)	5.7792 (0)
	(b)			3.7712 (0)	6.2363(−22)	6.0151 (9)	−3.9462 (0)	9.7532 (0)	5.8090 (0)
9	(a)	313.3325	−6.63002	1.284 (1)	2.504 (−21)	7.536 (9)	5.006 (−1)	1.7505 (0)	2.2511 (0)
	(b)			1.2826 (1)	2.5038(−21)	7.5248 (9)	4.9627(−1)	1.7583 (0)	2.2546 (0)
10	(a)	419.4304	−6.40000	4.435 (0)	8.6070(−21)	6.634 (8)	−2.3425 (0)	7.0961 (0)	4.7536 (0)
	(b)			4.1722 (0)	9.0028(−21)	5.7152 (8)	−2.7103 (0)	7.7300 (0)	5.0197 (0)
11	(a)	89.5143	−3.54002	4.23 (−1)	4.9275(−21)	3.271 (7)	−4.8293 (1)	9.7300 (1)	4.90072 (1)
	(b)			4.1815(−1)	4.9831(−21)	3.1584 (7)	−4.9194 (1)	9.9093 (1)	4.9899 (1)

[a] The numbers in parentheses indicate the powers of ten associated with the entries.

ACKNOWLEDGMENTS

The author expresses his appreciation to Dr. A. L. Latter and Dr. W. G. McMillan for helpful discussions and particularly to Mrs. F. Boyle, Miss R. Merrill, and the Rand Numerical Analysis Division for numerical assistance.

APPENDIX I

The effect of exchange in the temperature-dependent Thomas-Fermi model has been included and some solutions obtained by Askin[15] and by Cowan.[16] The formulation consists in modifying the density function of Eq. (7) by adding a term to the exponential in that equation which measures the exchange energy per electron at position $\mathbf{r}$ with momentum $\mathbf{p}$, namely

$$\rho(\mathbf{r},\mathbf{p}) = \frac{8\pi}{h^3} \frac{p^2}{\exp(-\alpha+p^2/2m-eV(r)+I(\mathbf{r},\mathbf{p}))/kT+1} \tag{A.1}$$

where $I(\mathbf{r},\mathbf{p})$ is the exchange energy. The quantity $I(\mathbf{r},\mathbf{p})$ may be expressed within the approximation of the Thomas-Fermi model by

$$I(\mathbf{r},\mathbf{p}) = \int d\mathbf{p}' V(\mathbf{p}-\mathbf{p}')\rho(\mathbf{r},\mathbf{p}')/p'^3 \tag{A.2}$$

where the integration extends over all momenta and $V(\mathbf{p}-\mathbf{p}')$ is the exchange potential energy between two plane-wave states of momenta $\mathbf{p}$ and $\mathbf{p}'$. The evaluation of $V(\mathbf{p}-\mathbf{p}')$ is quite straightforward and leads to

$$V(\mathbf{p}-\mathbf{p}') = -4\pi e^2\hbar^2/|\mathbf{p}-\mathbf{p}'|^2. \tag{A.3}$$

Substitution of Eq. (A.3) into Eq. (A.2) and carrying out that part of the integration over the directions of $\mathbf{p}'$ gives the result

$$I(\mathbf{r},p) = -\frac{e^2\hbar^2}{4\pi}\frac{1}{p}\int_0^\infty dp'\rho(\mathbf{r},p')\frac{1}{p'}\ln\left|\frac{p+p'}{p-p'}\right|. \tag{A.4}$$

Equations (A.1) and (A.4) define an integral equation for the evaluation of the exchange energy $I(\mathbf{r},p)$. The solution of this integral equation then determines completely the density function of Eq. (A.1), which in turn specifies the thermodynamic properties of the system by a simple modification of the arguments leading to Eqs. (10), (12), (15), and (31).

[15] J. Askin (unpublished).
[16] R. D. Cowan (unpublished).

Conservation of Charge in Einstein's Generalization of Gravitation Theory*

JAMES L. ANDERSON AND PARVIZ MERAT
Department of Physics, University of Maryland, College Park, Maryland
(Received April 14, 1955)

The consequences of λ invariance in Einstein's new modified field theory are investigated. It is shown that, as a consequence of this proposed invariance, it is possible to replace the antisymmetric part of the metric tensor with a four-potential. By further postulating invariance of the theory under a gauge transformation of the theory, it is possible to arrive at a quantity which can be interpreted as a current density four-vector.

INTRODUCTION

FROM our previous experiences, we have come to expect that theories which are invariant with respect to one or more continuous groups of transformations have associated with them a like number of conservation laws.[1] For instance, rotational invariance leads to the conservation law of angular momentum while gauge invariance of electromagnetic theory leads to the conservation of charge. It is important to note that the charge conservation law differs from the angular momentum conservation law in that the charge law is valid at each space-time point independent of whether the field equations are satisfied at this point while the angular momentum law depends upon the satisfaction of the field equations. This difference is intimately related to the nature of the related invariance group. The rotation group is a six-parameter Lie group while the gauge group is an infinite parameter Lie group; it is necessary to specify a function of the four-coordinates in order to specify a member of the group. The same situation pertains in general relativity. There it is necessary to specify four functions of the coordinates, the relations between the old and the new coordinates, in order to specify an element of the group. The corresponding conservation laws are those of energy and momentum.

In a revised version of his generalization of gravitation theory,[2] Einstein has proposed that his theory be invariant under a new type of invariance, the λ invariance. Under a λ transformation the Christoffel symbol $\Gamma^{\lambda}_{\mu\epsilon}$ transforms according to

$$\bar{\Gamma}^{\lambda}_{\mu\epsilon}=\Gamma^{\lambda}_{\mu\epsilon}+\delta^{\lambda}_{\mu}\lambda_{\epsilon}, \tag{1}$$

where δ^{λ}_{μ} is the Kronecker delta tensor and λ_{ϵ} is a set of four arbitrary space-time functions. Under this transformation, the Lagrangian

$$\mathfrak{L}=\mathfrak{g}^{\mu\epsilon}R_{\mu\epsilon}, \tag{2}$$

where

$$R_{\mu\epsilon}=[\Gamma^{\sigma}_{\mu\epsilon,\sigma}-\Gamma^{\lambda}_{\mu\sigma}\Gamma^{\sigma}_{\lambda\epsilon}]-[\Gamma^{\sigma}_{\mu\sigma,\epsilon}-\Gamma^{\lambda}_{\mu\epsilon}\Gamma^{\sigma}_{\lambda\sigma}] \tag{3}$$

changes by an amount

$$\begin{aligned}\delta\mathfrak{L}&=\mathfrak{g}^{\mu\epsilon}(\lambda_{\mu,\epsilon}-\lambda_{\epsilon,\mu})\\&=2\mathfrak{g}^{\mu\epsilon}\lambda_{\mu,\epsilon}.\end{aligned} \tag{4}$$

Here $\mathfrak{g}^{\mu\epsilon}$ represents the antisymmetric part of $\mathfrak{g}^{\mu\epsilon}$. The change in the integrated Lagrangian is thus

$$\begin{aligned}\delta\int\mathfrak{L}&=2\int\mathfrak{g}^{\mu\epsilon}\lambda_{\mu,\epsilon}d^4x\\&=-2\int\mathfrak{g}^{\mu\epsilon}{}_{,\epsilon}\lambda_{\mu}d^4x,\end{aligned} \tag{5}$$

provided of course that λ_{μ} vanish on the surface bounding the region of integration. Now in order that the theory be invariant under arbitrary λ transformations, the change in the integrated Lagrangian must vanish identically. Since the λ_{μ} are completely arbitrary we can conclude from Eq. (5) that

$$\mathfrak{g}^{\mu\epsilon}{}_{,\epsilon}\equiv 0. \tag{6}$$

These are the "Bianchi" identities associated with λ invariance. While it is true that they are already in the form of conservation laws, we cannot interpret them without further work.

FOUR-POTENTIALS

The Eqs. (6) are identities; they are valid regardless of any particular dependence of the field variables on their arguments. However, it is clear from the form of Eqs. (6) that we would encounter an inconsistency in the requirement of λ invariance if we interpreted the $\mathfrak{g}^{\mu\epsilon}$ as the fundamental field variables. The Eqs. (6) can only be satisfied by a very restricted set of field variables. On the other hand, if the $\mathfrak{g}^{\mu\epsilon}$ were themselves to depend upon other field variables, then it should be possible to arrange this dependence in such a manner that Eqs. (6) are really identities in these new field variables.

In order to find these new field variables and the dependence of $\mathfrak{g}^{\mu\epsilon}$ on them, we define the dual $\tilde{\mathfrak{g}}_{\mu\epsilon}$ to $\mathfrak{g}^{\mu\epsilon}$ by the equations

$$\tilde{\mathfrak{g}}_{\mu\epsilon}=\tfrac{1}{2}\eta_{\mu\epsilon\sigma\tau}\mathfrak{g}^{\sigma\tau}, \tag{7}$$

where $\eta_{\mu\epsilon\sigma\tau}$ is the Levi-Civita tensor density. Written

* This work was supported by the Office of Naval Research and the National Science Foundation.

[1] See E. L. Hill, Revs. Modern Phys. **23**, 253 (1951) for a discussion of the relation between invariance and conservation laws. For a treatment of the types of invariance and their conservation laws treated in this paper, see J. N. Goldberg, Phys. Rev. **89**, 263 (1953).

[2] A. Einstein, *The Meaning of Relativity* (Princeton University Press, Princeton, 1953), fourth edition, Appendix II.

in terms of $\tilde{\mathfrak{g}}_{\underset{\sim}{\mu\nu}}$, (6) becomes

$$(\tilde{\mathfrak{g}}_{\underset{\sim}{\rho\nu},\mu}+\tilde{\mathfrak{g}}_{\underset{\sim}{\mu\rho},\lambda}+\tilde{\mathfrak{g}}_{\underset{\sim}{\lambda\mu},\rho})=0, \tag{8}$$

from which it follows that $\tilde{\mathfrak{g}}_{\underset{\sim}{\mu\nu}}$ can be written as

$$\tilde{\mathfrak{g}}_{\underset{\sim}{\mu\nu}}=(\varphi_{\mu,\nu}-\varphi_{\nu,\mu}), \tag{9}$$

and thus we can write

$$\begin{aligned}\mathfrak{g}^{\underset{\sim}{\sigma}}&=\tfrac{1}{2}\eta^{\underset{\sim}{\sigma\mu\nu}}\tilde{\mathfrak{g}}_{\underset{\sim}{\mu\nu}}=\tfrac{1}{2}\eta^{\underset{\sim}{\sigma\mu\nu}}(\varphi_{\mu,\nu}-\varphi_{\nu,\mu})\\&=\eta^{\underset{\sim}{\sigma\lambda\mu}}\varphi_{\lambda,\mu}.\end{aligned} \tag{10}$$

It is therefore reasonable to take as our fundamental field variables the "potentials" φ_λ. In terms of them Eqs. (6) are indeed identities.

With these new variables the Lagrangian of the theory becomes

$$\mathfrak{L}=(\mathfrak{g}^{\underset{\sim}{\mu\nu}}+\eta^{\underset{\sim}{\nu\lambda\mu}}\varphi_{\lambda,\mu})R_{\underset{\sim}{\mu\nu}}, \tag{11}$$

and thus, considering the $\mathfrak{g}^{\underset{\sim}{\mu\nu}}$, φ_λ, and $\Gamma^\lambda{}_{\underset{\sim}{\mu\nu}}$ as independent field variables, our variational principal becomes

$$\int\{B^{\underset{\sim}{\mu\nu}}{}_\lambda\delta\Gamma^\lambda{}_{\underset{\sim}{\mu\nu}}+R_{\underset{\sim}{\mu\nu}}\delta\mathfrak{g}^{\underset{\sim}{\mu\nu}}+\eta^{\underset{\sim}{\nu\lambda\mu}}R_{\underset{\sim}{\mu\nu},\mu}\delta\varphi_\lambda\}d^4x=0. \tag{12}$$

Our field equations are then

$$B^{\underset{\sim}{\mu\nu}}{}_\lambda=-\mathfrak{g}^{\underset{\sim}{\mu\nu}}{}_{,\lambda}+\mathfrak{g}^{\underset{\sim}{\mu\nu}}{}_{,\sigma}\delta^\sigma{}_\lambda+\mathfrak{g}^{\underset{\sim}{\mu\nu}}\Gamma_\lambda+\mathfrak{g}^{\underset{\sim}{\sigma\nu}}\Gamma_\sigma\delta^\mu{}_\lambda=0, \tag{13a}$$

$$R_{\underset{\sim}{\mu\nu}}=0, \tag{13b}$$

$$\eta^{\underset{\sim}{\nu\lambda\mu}}R_{\underset{\sim}{\mu\nu},\mu}=0. \tag{13c}$$

(The notation is that of Einstein.) These equations are equivalent to those obtained by Einstein provided we choose λ_σ so that $\Gamma_\sigma=0$ with the exception that now φ_σ has replaced $\mathfrak{g}^{\underset{\sim}{\sigma\nu}}$ as a fundamental variable.

CURRENT DENSITY

As we mentioned in the Introduction, we can expect that a law of charge conservation will be associated with some invariance property of the theory akin to gauge invariance. The simplest way to introduce this invariance is to postulate that the theory is invariant under the potential transformation

$$\varphi_\lambda\rightarrow\varphi_\lambda+\Lambda_{,\lambda}, \tag{14}$$

where Λ is an arbitrary function of the space-time coordinates.

In order that the Lagrangian (11) be invariant under this transformation, it is sufficient to require that

$$\eta^{\underset{\sim}{\nu\lambda\mu}}R_{\underset{\sim}{\mu\nu},\mu\lambda}\equiv 0 \tag{15}$$

as can be seen from Eq. (12). Because of the symmetry properties of $\eta^{\underset{\sim}{\nu\lambda\mu}}$ it is evident that the left-hand-side of (15) is indeed identically zero. It does not seem unreasonable therefore, to take as our current-density four-vector

$$\mathfrak{I}^\lambda=\eta^{\underset{\sim}{\nu\lambda\mu}}R_{\underset{\sim}{\mu\nu},\mu}, \tag{16}$$

with the associated conservation law

$$\mathfrak{I}^\lambda{}_{,\lambda}=0. \tag{17}$$

The total charge enclosed in a region Ω is given by

$$Q=\int_\Omega\eta^{\underset{\sim}{\nu\lambda\mu}}R_{\underset{\sim}{\mu\nu},\mu}d^4x, \tag{18}$$

which, by Gauss' theorem can be written as

$$Q=\int_\Sigma\eta^{\underset{\sim}{\nu\lambda r}}R_{\underset{\sim}{\mu\nu}}d\sigma_r, \tag{18}$$

where $d\sigma_r$ is a surface element with normal along the r coordinate axis. It is evident from Eq. (17) that, for a charge concentrated within a finite volume, the integral is independent of the surface Σ provided only that it completely enclosed the charge. We should require this property if the integral is to represent the charge enclosed within a given region of space.

DISCUSSION

In order to obtain any information at all from a theory like the one we have been considering, it is necessary to construct quantities like the energy-momentum tensor and the current-density vector in order that a connection can be made with experiment. The field equations by themselves tell us nothing since in general we do not know how to interpret the field variables which appear therein. In order that quantities which represent the physical observables of the theory are not introduced in a completely ad hoc manner, we look for some guiding principle to aid us in our choice. At present that principle seems to lie in the close connection between invariances of the theory and the corresponding conservation laws. We have used this principle in postulating that the right-hand side of Eq. (16) really does represent the current density of the theory.

Our assertion differs from that made by Einstein. He postulates that

$$\mathfrak{I}^\mu=\eta^{\underset{\sim}{\nu\lambda\mu}}g_{\underset{\sim}{\nu},\lambda}, \tag{20}$$

and indeed it is true that

$$\mathfrak{I}^\mu{}_{,\mu}=\eta^{\underset{\sim}{\nu\lambda\mu}}g_{\underset{\sim}{\nu},\lambda\mu}\equiv 0, \tag{21}$$

so that the quantity appearing on the right-hand side of (20) is conserved. However, it is not at all evident that there are not many other quantities which possess this property and indeed we have found one such additional quantity. We believe that one must have a choice based on a general principle applicable to all cases, such as the aforementioned relation between invariance and conservation laws.

Our choice (16) however is not completely free from objection. Einstein has pointed out[2] that our current-

[2] A. Einstein (private communication).

density vanishes everywhere that the field equations are satisfied and hence is unsatisfactory since the only acceptable solutions to the field equations are those which satisfy them everywhere. Thus, for this solution, our current density would be everywhere zero. We believe, however, that it is necessary to consider solutions of the field equations which do not satisfy them everywhere. In certain small regions of space it might well be that the field equations are not even valid. One can take the position that the field equations describe correctly the interaction of elementary particles but are insufficient to describe the particles themselves and hence, at small distances from these particles, must be replaced by some other kind of mathematical construction.[4] If such were the case we would no longer need to restrict our solutions and the current density of Eq. (16) could be nonzero for certain regions of space. In fact, if one takes seriously the relation between invariance and conservation laws, one is almost forced into this latter position. Regardless of the form of the Lagrangian, the current density which follows from gauge invariance will vanish whenever the field equations are satisfied.

The only way out seems to be to construct a theory in which it is impossible to introduce potentials. While there would still be an invariance associated with the conservation of charge, it may be of such a nature that the current-density need not vanish when the field equations are satisfied. For instance, the energy-momentum tensor associated with arbitrary coordinate invariance does not vanish when the field equations are satisfied. It is evident that a theory for which it is impossible to introduce potentials will differ from the present theory and hence lies outside the scope of our investigation.

[4] This view is diametrically opposite to that held by Einstein. He contends that his field equations are everywhere valid, i.e., that they do correctly describe the elementary particles.

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Gravitational Radiation

JOSHUA N. GOLDBERG
Armour Research Foundation, Illinois Institute of Technology, Chicago, Illinois
(Received October 19, 1953; revised manuscript received May 27, 1955)

In this paper, we investigate the existence of gravitational radiation within the framework of the EIH approximation method. Following a prescription introduced by Infeld, the radiation terms of the EIH expansion are begun with functions of the time alone. We find that these terms do not have physical significance if they are introduced in the scalar or longitudinal components of the gravitational potentials. However, if the radiation terms are introduced in the fifth order of the transverse-transverse components, one finds a contribution to the curvature tensor in the seventh order, a contribution to the equations of motion in the ninth order, and radiation in the tenth order. The existence of radiation is determined by calculating the energy passing through a spherical surface which is an infinite distance from all source points. This definition of radiation agrees with that used in the theory of electromagnetism.

I. INTRODUCTION

IN recent years there has been some controversy concerning gravitational radiation. Infeld and Scheidegger have maintained that the possibility of radiation does not exist within the framework of the EIH (Einstein, Infeld, and Hoffman) approximation method.[1-4] This conclusion has been accepted for the scalar and longitudinal components. Indeed, in his book Bergmann has shown[5] that in the linearized gravitational equations these terms do not contribute to radiation; however, the transverse-transverse components do make a contribution. It was on this basis that the proof set forth by Infeld and Scheidegger was first criticized.[6] This paper will attempt to clarify the situation by proposing an unambiguous definition of radiation.

II. EIH APPROXIMATION METHOD

In a previous paper[7] the surface integrals leading to the equations of motion were found without recourse to an approximation method. The possibility of doing so depended on the existence of superpotentials for the components of the energy-momentum psuedo-tensor:*

$$T_\mu{}^\nu = U_\mu{}^{[\nu\sigma]}{}_{,\sigma}. \tag{1}$$

$T_\mu{}^\nu$ is the energy-momentum psuedo-tensor and $U_\mu{}^{[\nu\sigma]}$

[1] L. Infeld and A. E. Scheidegger, Can. J. Math. **3**, 195 (1951).
[2] A. E. Scheidegger, Phys. Rev. **82**, 883 (1951).
[3] L. Infeld, Can. J. Math. **5**, 17 (1953).
[4] A. E. Scheidegger, Revs. Modern Phys. **25**, 451 (1953).
[5] P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall Publications, Inc., New York, 1947), pp. 187–189.
[6] P. G. Bergmann (private communication).
[7] J. N. Goldberg, Phys. Rev. **89**, 263 (1953).
* The energy-momentum pseudo-tensor does not have simple geometrical transformation properties (see reference 5, page 196). Hence, the word *pseudo-tensor* should not be confused by the current use of the prefix *pseudo-* to describe a density of weight one.

is the antisymmetric superpotential. As a result, it was shown that whenever the field equations of a covariant field theory are satisfied on a two-dimensional closed surface, that is,

$$L^A=0 \tag{2}$$

on the two-dimensional surface, the following integrals taken over that surface vanish:

$$\Sigma_\mu=\oint (U_\mu^{[\nu\sigma]}{}_{,\sigma}+t_\mu^{\ \nu})n_\nu dS=0. \tag{3}$$

In the above equations, L^A are the field equations and $t_\mu^{\ \nu}$ is equal to the energy-momentum psuedo-tensor when the field equations are satisfied:

$$t_\mu^{\ \nu}=-\delta_\mu^{\ \nu}L+y_{A,\mu}\partial^{A\nu}L. \tag{4}$$

The above surface integrals, Eq. (3), were shown to be equivalent to those used by EIH in their solution of the problem of motion in the general theory of relativity. They are independent of the surface chosen in the sense that the restrictions placed on the coordinates of any singularities enclosed by the surface are the same regardless of the choice of surface. For our purposes it will be convenient to discuss the EIH approximation method using the notation of this paragraph.

The method of EIH[8–10] assumes that the particles generating the gravitational field are represented by singularities in the field and that the velocities of these particles are small compared with that of light ($v/c\ll 1$). It follows that the derivative of the field variables (gravitational potentials) with respect to time will be of a higher order in v/c than the derivatives with respect to the space coordinates. One introduces a parameter τ such that $\tau=\lambda x^4(\lambda=v/c)$. Derivatives taken with respect to τ are assumed to be of the same order as those with respect to the space coordinates. τ differentiation will be represented by a zero following a comma; i.e., $y_{A,4}=\lambda y_{A,0}$. For the sake of consistency, in the following Greek indices will run from 0 to 3. This notation has the advantage of constantly reminding us of the nature of the approximation method. As usual, Latin indices run from 1 to 3. Finally, the field variables, y_A, are expanded into a power series in λ:

$$y_A=\sum_{n=0}^{\infty}\lambda^n{}_n y_A. \tag{5}$$

As a result, the field equations and the surface integrals will also be power series in λ and the coefficients of the various powers of λ may be equated to zero separately:

$${}_nL^A=0, \tag{6}$$

$${}_n\Sigma_\mu=0. \tag{7}$$

Let us assume that the field equations have been solved and the surface integrals satisfied through the nth order. Therefore we know ${}_0y_A$, $\cdots$, ${}_ny_A$. The superpotential ${}_nU_\mu^{[\nu\sigma]}$ contains terms which are linear in ${}_ny_A$ as well as the lower order field variables in combinations of higher degree. In the surface integrals, however, this term is differentiated with respect to τ and therefore is of order $n+1$. The components of the weak energy-momentum psuedo-tensor $t_\mu^{\ \nu}$ are homogeneous quadratic in the first derivatives of the field variables and therefore knowledge of the field variables to the nth order gives us $t_\mu^{\ \nu}$ at least to the $(n+1)$th order. Therefore, if the solution of the field equations up to the nth order is to be consistent, the surface integrals in the $(n+1)$th order must vanish. In general one cannot expect these surface integrals to be satisfied when the surface encloses a singularity. However, Einstein and Infeld have shown[10] that in the theory of gravitation ${}_{n+1}\Sigma_4$ may be satisfied by the addition of poles in the nth order and ${}_{n+1}\Sigma_s$ by the addition of dipoles in the $(n-1)$th order. When the solution has been carried as far as desired, the sum of all the dipoles added is set equal to zero. That this process yields the equations of motion for the singularities follows from the fact that the surface integrals are merely conservation laws for linear momentum. For arbitrary particle motions, linear momentum can be conserved by the addition of appropriate time dependent dipoles. The ultimate prohibition of such dipoles forces the particles to move in such a manner as to conserve linear momentum.

There is still a certain amount of arbitrariness in the solution and hence in the equations of motion which has not been discussed. In the nth order the field equations will consist of linear terms involving the field variables of the nth order and linear as well as nonlinear terms involving the field variables of lower orders. Thus in each order, except for the lowest, one has to solve inhomogeneous linear partial differential equations. Therefore, to the solution of the inhomogeneous equations one can add an arbitrary solution of the homogeneous equations. One must decide in advance what singularities are to be permitted if unique equations of motion are to be obtained. Having chosen a particular solution in the lowest order, EIH prohibit the addition of any other arbitrary solutions of the homogeneous equations.

Apart from the above, there is an additional arbitrariness which results from the question of radiation: Should one choose a retarded, advanced, or a linear combination of retarded and advanced solutions? EIH avoid the question of radiation by choosing the standing wave solution (retarded plus advanced).

In the theory of gravitation it is convenient to introduce linear combinations of the deviation from flat space as the field variables in the approximation method:

$$\gamma_{\mu\nu}=h_{\mu\nu}-\tfrac{1}{2}\eta_{\mu\nu}\eta^{\sigma\tau}h_{\sigma\tau}, \tag{8}$$

[8] Einstein, Infeld, and Hoffman, Ann. Math **39**, 66 (1938).
[9] A. Einstein and L. Infeld, Ann. Math **41**, 455 (1940).
[10] A. Einstein and L. Infeld, Can. J. Math **1**, 209 (1949).

where

$$h_{\mu\nu}=g_{\mu\nu}-\eta_{\mu\nu} \tag{9}$$

is the deviation from flat space. $\eta_{\mu\nu}$ $(-1,-1,-1,+1)$ is the Minkowski metric. The $\gamma_{\mu\nu}$ are then expanded into a power series in λ beginning with λ^2:

$$\begin{aligned}\gamma_{00}&=\sum_{n=1}^{\infty}\lambda^{2n}\,{}_{2n}\gamma_{00}+\sum_{n=1}^{\infty}\lambda^{2n+1}\,{}_{2n+1}\gamma_{00},\\ \gamma_{0r}&=\sum_{n=1}^{\infty}\lambda^{2n+1}\,{}_{2n+1}\gamma_{0r}+\sum_{n=1}^{\infty}\lambda^{2n+2}\,{}_{2n+2}\gamma_{0r},\\ \gamma_{rs}&=\sum_{n=1}^{\infty}\lambda^{2n+2}\,{}_{2n+2}\gamma_{rs}+\sum_{n=1}^{\infty}\lambda^{2n+3}\,{}_{2n+3}\gamma_{rs}.\end{aligned} \tag{10}$$

As indicated above, the expansions of the $\gamma_{\mu\nu}$ separate into two independent sets. The field equations have the property that each set must be introduced separately. By comparing the above expansion with the corresponding expansion of the electromagnetic theory, Infeld[11] concluded that the first set represents an advanced plus retarded field (standing waves) while the second set represents a retarded minus advanced field. Therefore, he called the second group of terms the "radiation terms." The solution of EIH considers only the standing waves.

In the theory of electromagnetism, the radiation terms start off with a function of time alone. Arguing by analogy, Infeld assumed that this situation should also be true in the theory of gravitation. Together with Scheidegger, he has shown that any function of time introduced into the radiation terms can be eliminated by means of a coordinate transformation.[1] Therefore, they concluded that gravitational radiation does not exist in the EIH framework.

The fallacy of their argument can be shown at this point. Let us consider the curvature tensor (see Appendix 2). The linear terms of this tensor contain only second derivatives of the field variables while higher degree terms contain a part which is homogeneous-linear in the second derivatives and a part which is homogeneous-quadratic in the first derivatives of the field variables. It is clear then that the introduction of an arbitrary function of time into the nth order of the radiation terms may make a contribution to the curvature tensor not in the nth order, but in the $(n+2)$th order. Thus, to say that the arbitrary function of time which has been introduced may be transformed away in the nth order is trivial, for it does not contribute to anything physical in that order. It follows, then, that one must investigate the higher order terms in order to determine whether anything physical results from the introduction of a function of time into the radiation terms.

This discussion may be clarified by examining the corresponding situation in the theory of electromagnetism where we know what the physically meaningful variables are. In this case the field equations for the radiation terms of the vector potential are

$${}_{2n+1}\varphi_{0,\,ss}={}_{2n-1}\varphi_{0,\,00},\quad {}_{2n+2}\varphi_{r,\,ss}={}_{2n}\varphi_{r,\,00}, \tag{11}$$

with the gauge condition

$${}_{2n+1}\varphi_{0,\,0}-{}_{2n+2}\varphi_{r,\,r}=0. \tag{12}$$

If we choose to begin the radiation terms with

$${}_{2n}\varphi_{r}=f_r(\tau), \tag{13}$$

this term may be removed by the gauge transformation generated by

$${}_{2n}\psi=-x^n f_n. \tag{14}$$

However, ${}_{2n}\varphi_r$ does not contribute to any physical quantity in the $(2n)$th order, but rather in the $(2n+1)$th order. Choosing

$${}_{2n+1}\varphi_0=0, \tag{15}$$

we find for the electric field strength

$${}_{2n+1}E_r={}_{2n+1}\varphi_{0,\,r}-{}_{2n}\varphi_{r,\,0}=-f_{r,\,0}. \tag{16}$$

Clearly the gauge transformation of Eq. (14) which removes ${}_{2n}\varphi_r$ introduces

$${}_{2n+1}\varphi_0=-x^n f_{n,\,0}, \tag{17}$$

so that Eq. (16) is maintained as it should be. Had the higher order effect of the gauge transformation been neglected, we would no longer have been considering the same physical problem we started with.

In our investigation, the coordinate transformations will be used to simplify the necessary calculations.

III. DEFINITION OF RADIATION

Thus far nothing has been said about how one is to know whether or not radiation exists. Certainly the existence of so-called "radiation" terms does not prove that radiation occurs. In the theory of electromagnetism one calculates the energy flux through a closed two-dimensional surface. If there is a finite flux through the surface when the surface is infinitely far from all source points, one says that radiation occurs. The existence of the strong conservations laws in the general theory of relativity allows the same definition to be applied in this case. The energy and momentum density of the gravitational field, including the source points, is given by $(1/16\pi\kappa)T_\mu{}^\nu$ where κ is the gravitational constant. Thus, the total energy and momentum, P_μ, contained in a volume V with a surface S is given by

$$P_\mu=\frac{1}{16\pi\kappa}\int_V T_\mu{}^\nu dV. \tag{18}$$

The negative time derivative of P_μ yields the flow of

[11] L. Infeld, Phys. Rev. 53, 836 (1938).

energy and momentum, W_μ, out of the volume V:

$$W_\mu = -\frac{1}{16\pi\kappa}\int_V T_\mu^0{}_{,0}\,dV = \frac{1}{16\pi\kappa}\oint T_\mu^\nu\, n_\nu dS. \tag{19}$$

In the last step, we have made use of the strong conservation laws

$$T_\mu^\nu{}_{,\nu} \equiv 0,$$

and then applied Gauss' theorem. If the field equations are satisfied on the surface S, the weak energy-momentum tensor, Eq. (4) may be substituted into Eq. (19):

$$W_\mu = \frac{1}{16\pi\kappa}\oint t_\mu^\nu\, n_\nu dS. \tag{20}$$

We shall say that gravitational radiation exists if the surface integral of Eq. (20) with $\mu=0$ yields a finite result when the surface is infinitely far from all source points. The total flow of momentum may be similarly defined with $\mu=1, 2, 3$.

By means of the surface integrals of Eq. (3), or directly from the definition of the superpotentials in Eq. (1), the radiation may be defined in terms of the superpotentials. However, as pointed out previously, the superpotentials contain linear terms whereas the t_μ^ν do not. Therefore, once the field equations have been solved up to a given order, Eq. (20) permits calculation of the radiation to a higher order than does the corresponding surface integral involving the superpotentials.

One can easily show, though we shall not do so here, that with the above definition of radiation the solution obtained by EIH does not contain radiation and hence corresponds to a standing wave.

IV. TRANSFORMATION EQUATIONS

We shall be concerned with coordinate transformations which change the coordinate values by an amount proportional to a given power of λ:

$$x^\rho = x'^\rho + \lambda^m\, {}_m\xi^\rho(x'). \tag{21}$$

In order to preserve the slow time variation of the field variables, the transformation functions should also possess this character. From the transformation properties of the gravitational potentials, the g_μ^ν, one can establish that the $\gamma_{\mu\nu}$ transform as follows:

$$\begin{aligned}
&\text{(a)} && {}_m\gamma'_{00} = {}_m\gamma_{00} - {}_m\xi^n{}_{,n},\\
&\text{(b)} && {}_m\gamma'_{0r} = {}_m\gamma_{0r} + {}_m\xi^0{}_{,r},\\
&\text{(c)} && {}_m\gamma'_{rs} = {}_m\gamma_{rs} - {}_m\xi^r{}_{,s} - {}_m\xi^s{}_{,r} + \delta_{rs}\, {}_m\xi^n{}_{,n},\\
&\text{(d)} && {}_{m+1}\gamma'_{00} = {}_{m+1}\gamma_{00} + {}_m\xi^0{}_{,0},\\
&\text{(e)} && {}_{m+1}\gamma'_{0r} = {}_{m+1}\gamma_{0r} - {}_m\xi^r{}_{,0},\\
&\text{(f)} && {}_{m+1}\gamma'_{rs} = {}_{m+1}\gamma_{rs} + \delta_{rs}\, {}_m\xi^0{}_{,0}.
\end{aligned} \tag{22}$$

In the above equations all functions are considered to be functions of the new coordinates and all derivatives are taken with respect to the new coordinates.

There are other higher order effects of the transformation which we shall not consider. These higher order terms will combine with the untransformed solution to produce solutions either of the homogeneous or inhomogeneous equations in the higher orders. Since the solution of the inhomogeneous equations will be generated as we proceed with the approximation method, we lose only solutions of the homogeneous equations by disregarding higher order terms in the transformation equations. In view of the previously discussed ambiguity of the solution all such solutions of the homogeneous equations may be discarded. It is clear from the above discussion that the prohibition of solutions of the homogeneous equations is not a covariant requirement. Beyond the lowest nonvanishing order we shall always prohibit the appearance of arbitrary solutions of the homogeneous equations in that coordinate system in which we are working. Whether or not there exist other coordinate systems, in which this condition is also satisfied, will not concern us.

We are now prepared to solve the field equations for the radiation terms. The fundamental problem to be considered is that of two point masses interacting. EIH have already solved the stationary wave part of this problem. We shall look for a nonsingular solution of the field equations (radiation terms) which cannot be removed by a coordinate transformation. Having found such a solution, we must then investigate, in the manner of Sec. III, whether or not radiation of energy occurs. The introduction of radiation terms will not alter the EIH solution. Fortunately, for our purposes we shall only require ${}_2\gamma_{00}$. As has been mentioned previously, we do not question Infeld and Scheidegger's result if the radiation terms are begun in the scalar or longitudinal components. However, for the sake of completeness we shall consider these cases as well as that of the transverse-transverse components.

In order not to obscure the principal argument with long cumbersome equations, we shall put the gravitational field equations into an appendix and bring forward only those terms required for our purposes. Equations in the appendices will be referred to as Eq. (A7), etc.

V. SCALAR AND LONGITUDINAL COMPONENTS

Although the radiation terms may be introduced in any order, we shall do so as early as possible, i.e., into ${}_2\gamma_{00}$, ${}_4\gamma_{0r}$, and ${}_5\gamma_{rs}$. The argument is not confined to this choice, for the same types of terms are combined in the field equations and the surface integrals regardless of the order. The field equations in any order may be found from Appendix 1. In all our considerations we shall begin by assuming the standard coordinate

conditions introduced by EIH:

$$ {}_m\gamma_{00,0} - {}_{m+1}\gamma_{0r,r} = 0, \tag{23} $$

$$ {}_m\gamma_{rs,s} = 0. \tag{24} $$

However, we shall not restrict the coordinate transformations to those which preserve these conditions.

The field equations for the scalar and longitudinal components, in the third and fourth orders respectively, are

$$ {}_3\gamma_{00,ss} = 0, \tag{25} $$

$$ {}_4\gamma_{0r,ss} = 0. \tag{26} $$

In the above equations, the coordinate conditions have already been introduced. Following the prescription given by Infeld,[11] we start the scalar term with the solution

$$ {}_3\gamma_{00} = f(\tau). \tag{27} $$

Before proceeding to the next stage of the approximation, the surface integral ${}_4\sum_0$ [Eq. (3)] should be evaluated to insure the consistency of the solution ${}_3\gamma_{00}$. Since the value of the surface integral is independent of the specific surface of integration, the only terms of the integrand which can lead to a nonvanishing integral are those which have an r^{-2} behavior. Clearly, such terms cannot appear in ${}_4\sum_0$ and as a result the consistency of the solution in the third order is assured. This argument will be valid throughout our discussion as we proceed from the solution in one order to that in the next. Therefore, in all such cases we need not be concerned with the surface integrals.

The solution for ${}_4\gamma_{0r}$ which is consistent with the coordinate conditions is given by [Eqs. (26) and (23)]

$$ {}_4\gamma_{0r} = \tfrac{1}{4}x^r f_{,0}. \tag{28} $$

Now another question arises in relation to the surface integrals: Is it necessary to examine ${}_3\gamma_{00}$ and ${}_4\gamma_{0r}$ for consistency before carrying out a coordinate transformation to remove ${}_3\gamma_{00}$? More generally the question is whether it is necessary to examine the consistency of a solution before carrying out a coordinate transformation which affects that solution. If the solution is not consistent, then appropriate poles and dipoles must be added to it. After carrying out the transformation, we must examine the transformed solutions for consistency. The net result will be to alter the poles and dipoles which had been added. Therefore, we can delay investigating the surface integrals in a given order until a coordinate transformation has been carried out which affects the solution in a lower order only. As a result of this and the previous argument we shall not need to consider the surface integrals until the solutions have been carried as far as desired.

A coordinate transformation may now be carried out to remove ${}_3\gamma_{00}$. From Eq. (22), with

$$ {}_3v^r = \tfrac{1}{4}x^r f(\tau), \tag{29} $$

and

$$ {}_3v^0 = -\tfrac{1}{4}\int f(\tau)\,d\tau, $$

we find

$$ {}_3\gamma'_{00} = 0, \tag{30a} $$

and

$$ {}_4\gamma'_{0r} = \tfrac{1}{12}x^r f_{,00}. \tag{30b} $$

Clearly the coordinate conditions in the fourth order have been altered by the transformation. In the fifth order, the field equations are

$$ {}_5\gamma_{rs,nn} - {}_5\gamma_{rn,sn} - {}_5\gamma_{sn,rn} + \delta_{rs}\,{}_5\gamma_{nn,mm} = -{}_4\gamma_{0r,0s} - {}_4\gamma_{0s,0r} + 2\delta_{rs}\,{}_4\gamma_{0n,0n} - \delta_{rs}\,{}_5\gamma_{00,00} \tag{31a} $$

$$ {}_5\gamma_{00,nn} = {}_5\gamma_{mn,mn}, \tag{31b} $$

and the coordinate conditions are

$$ {}_5\gamma_{rs,s} = 0. \tag{31c} $$

Hence,

$$ {}_5\gamma'_{rs,nn} = \tfrac{1}{3}\delta_{rs}f_{,00}, \tag{32a} $$

$$ {}_5\gamma'_{00,nn} = 0. \tag{32b} $$

The solutions of these equations may be found in a straightforward manner:

$$ {}_5\gamma'_{rs} = (1/15)\delta_{rs}r^2 f_{,00} - (1/30)x^r x^s f_{,00}, \tag{33a} $$

$$ {}_5\gamma'_{00} = 0. \tag{33b} $$

Another coordinate transformation may be carried out to eliminate the longitudinal component in the fourth order. Consider

$$ {}_4v^0 = -(1/24)r^2 f_{,0}. \tag{34} $$

From Eqs. (22), we find that

$$ {}_5\gamma''_{rs} = [(1/40)\delta_{rs}r^2 - (1/30)x^r x^s]f_{,00}, \tag{35a} $$

$$ {}_5\gamma''_{00} = -(1/24)r^2 f_{,00}. \tag{35b} $$

The field equations in the sixth order are

$$ {}_6\gamma_{0s,nn} - {}_6\gamma_{0n,ns} = {}_5\gamma_{rs,0r} - {}_5\gamma_{00,0s}, \tag{36a} $$

with the coordinate condition

$$ {}_6\gamma_{0s,s} - {}_6\gamma_{00,0} = 0. \tag{36b} $$

From Eqs. (35), we obtain

$$ {}_6\gamma''_{0s,nn} = -(1/12)x^s f_{,000}, \tag{37a} $$

$$ {}_6\gamma''_{0s,s} = -(1/24)r^2 f_{,000}. \tag{37b} $$

The solution of these equations is

$$ {}_6\gamma''_{0r} = -(1/120)x^r r^2 f_{,000}. \tag{38} $$

Now, one final transformation may be carried out which wipes out the entire solution. An examination

of the transformation equations shows that

$$ {}_5v^s = -(1/120)x^s r^2 f_{,00} \tag{39} $$

is the desired transformation function.

Now let us examine what happens if the radiations terms are begun in the longitudinal components with the solution

$$ {}_4\gamma_{0r} = f_r(\tau). \tag{40} $$

In this case the proof is much simpler for we need not consider the higher order solutions. Let

$$ {}_2v^r = \int f_r(\tau)d\tau. \tag{41} $$

An examination of Eqs. (22) shows that ${}_4\gamma'_{0r}$ vanishes and that no other term is affected. Thus, the entire solution is wiped out by the coordinate transformation.

One may argue that if higher order effects of the coordinate transformations had not been neglected, the solution as a whole could not have been removed by the transformations considered. Furthermore, by our stated intent to keep solutions of the homogeneous equations in the lowest nonvanishing order, we should have been obliged to consider the higher order effects. In the lowest orders the solutions introduced by these higher order effects could have been of two types: nonsingular functions of time alone and singular functions with time dependent coefficients. The purely time dependent functions could be removed by coordinate transformations and hence are of no consequence. The singular terms, on the other hand, would correspond to time-dependent multipoles. Such terms cannot be of physical importance because we always have the possibility of altering the equations of motion by the addition of multipoles.

VI. TRANSVERSE-TRANSVERSE COMPONENTS

The field equations for γ_{rs} in the fifth order are

$$ {}_5\gamma_{rs,nn} = 0, \tag{42} $$

if the coordinate conditions given in Eqs. (23) and (24) are used. Therefore, in this case the solutions may also be started off with an arbitrary function of time:

$$ {}_5\gamma_{rs} = f_{rs}(\tau). \tag{43} $$

The transformations to be considered in this case are handled in the same manner as those in the previous section. Therefore, in order to avoid the monotony of repetition, we shall merely state the results of the transformations and relegate the explicit treatment of the coordinate transformations to the Appendix (Appendices 4 and 5).

Starting from an arbitrary, though symmetric, set of functions $f_{rs}{}^0(\tau)$ one can always transform the set to another, $f_{rs}(\tau)$, with vanishing trace (Appendix 4); i.e., such that

$$ f_{ss}(\tau) = 0. \tag{44} $$

This result is in agreement with the discussion of gravitational radiation given in Bergmann's book.[5] Continuing the solution with the above condition, we find, after the appropriate transformations, that the lowest nonvanishing order is the seventh (Appendix 5). In the seventh and eighth orders the solutions are

$$ {}_7\gamma_{rs} = (12/77)r^2 f_{rs,00} - (18/77)x^n(x^r f_{ns} + x^s f_{nr})_{,00} + (3/7)\delta_{rs}x^m x^n f_{mn,00}, \tag{45a} $$

$$ {}_7\gamma_{00} = 0, \tag{45b} $$

$$ {}_8\gamma_{0r} = 0. \tag{45c} $$

The standard coordinate conditions are satisfied by these solutions.

The question still remains whether the above solution leads to anything physical. An examination of the curvature tensor is the simplest test of this point. From Eq. (A15) we find that the $(s00,r)$ component of the curvature tensor in the seventh order is

$$ {}_7R_{s00}{}^r = -\tfrac{1}{4}({}_7\gamma_{00,rs} + {}_7\gamma_{nn,rs}). \tag{46} $$

Substituting Eqs. (35), we have

$$ {}_7R_{s00}{}^r = -(9/22)f_{rs,00}. \tag{47} $$

Thus, it is clear that introducing the radiation terms through the transverse-transverse components has a physical significance.

Before computing the radiation in the tenth order, we shall examine the surface integrals in the ninth order, ${}_9\sum_m$, to establish the consistency of our solution and also because there is a contribution to the equations of motion in this order. From Eqs. (A9) and (A10), we find

$$ {}_9\sum_m = \oint [-{}_7\gamma_{ms,00} - \tfrac{1}{4}\delta_{sm}\,{}_2\gamma_{00,r}\,{}_7\gamma_{nn,r} + \tfrac{1}{4}{}_2\gamma_{00,n}\,{}_7\gamma_{nn,s} + \tfrac{1}{4}{}_2\gamma_{00,s}\,{}_7\gamma_{nn,m}]n_s dS. \tag{48} $$

Previously, we stated that the problem we were considering involved two particles interacting. Up to this time we have not required this information. The solution of ${}_2\gamma_{00}$ involving two particles is[4,10]

$$ {}_2\gamma_{00} = -4(m_1/r_1 + m_2/r_2), \quad (r_1)^2 = (x^s - \xi^s)(x^s - \xi^s), \quad (r_2)^2 = (x^s - \eta^s)(x^s - \eta^s), \tag{49} $$

ξ^s and η^s specify the locations of the particles. Inserting Eqs. (45a) and (49) into (48) and taking a surface which encloses only the first particle, we find the following contribution to the equations of motion:

$$ (1/4\pi)_9\sum_m = (18/11)m_1\xi^n f_{nm,00}. \tag{50} $$

The surface integrals have been computed with the aid

of Appendix 3. One cannot say more about this contribution to the equations of motion until information is available as to the form of f_{rs}.

If the momentum passing through a surface at infinity which encloses both particles is now computed, [Eq. (20)], one obtains

$$_9W_m=(9/22\kappa)(m_1\xi^n+m_2\eta^n)f_{nm,00}. \tag{51}$$

This momentum transfer vanishes if the origin is taken at the center of mass of the system. Since the momentum transfer is of a higher order in v/c than the energy transfer, there can be no passage of momentum through a surface at infinity in this order. This result implies that the radiation appears to come from the center of mass of the system. Indeed, if in the solution for $_7\gamma_{rs}$, $\mathbf{x}$ is replaced by $\mathbf{x}-\mathbf{x}'$, where $\mathbf{x}'$ is the center of mass, the momentum transfer vanishes regardless of the location of the origin. For the sake of simplicity we shall assume the origin to be at the center of mass.

Because of the quadratic nature of the $t_\mu{}^\nu$, solutions of higher order than the eighth cannot contribute to the radiation in the tenth order. Also, we shall not add the dipole to $_7\gamma_{00}$ as this term cannot contribute to the radiation because it falls off too rapidly as a function of r. Therefore, the only terms we need be concerned with are $_7\gamma_{rs}$ [Eq. (45a)] and $_2\gamma_{00}$ [Eq. (49)]. We find from Eq. (A9) that these terms make the following contribution to $t_0{}^s$:

$$_{10}t_0{}^s=-\tfrac{1}{4}(_2\gamma_{00,0}\,{}_7\gamma_{nn,s}+{}_2\gamma_{00,s}\,{}_7\gamma_{nn,0}). \tag{52}$$

If the above expression is now integrated over an infinite spherical surface centered at the origin, we obtain, with the aid of Appendix 3,

$$_{10}W_0=(9/55\kappa)[(m_1\xi^r\xi^s{}_{,0}+m_2\eta^r\eta^s{}_{,0})f_{rs,00}-\tfrac{3}{4}f_{rs,000}(m_1\xi^r\xi^s+m_2\eta^r\eta^s)]. \tag{53}$$

Thus, the introduction of the radiation terms through the transverse-transverse components in the fifth order lead to gravitational radiation in the tenth order and thereby justifies the nomenclature.

There is one other point which requires discussion. In a recent paper,[8] Infeld has shown that one can eliminate the contribution to the equations of motion in all orders higher than the sixth by an appropriate choice of coordinate conditions. The equations of motion through the sixth order are independent of the coordinate conditions.[10] However, in his proof he does not consider the radiation terms. By the same method, one can prove for the radiation terms that the contribution to the equations of motion through the ninth order does not depend on the coordinate conditions in the seventh order. Indeed, one can go further and prove that the radiation in the tenth order also does not depend on the coordinate conditions in the seventh order. Therefore, gravitational radiation has an invariant significance.

VII. CONCLUSION

We have examined the effects of introducing functions of time into the fifth order of the radiation terms and have found that they lead to a finite curvature tensor in the seventh order, a contribution to the equations of motion in the ninth order, and gravitational radiation in the tenth order. However, these results in no way constitute a proof of the existence of gravitational radiation for a detailed discussion of these effects cannot be given without knowing the specific form of the functions f_{rs} which were introduced in the fifth order. Clearly the choice of the f_{rs} will depend on what radiation effects are sought—those resulting from an externally applied field or from the interaction of two or more mass points. If we restrict our attention to the latter case, which is the case of principal interest, the f_{rs} are no longer arbitrary. Once we decide to consider a retarded (or advanced) solution, the starting functions in the radiation terms are related to the standing wave solutions in a precise manner. For example, consider the retarded and advanced potentials

$$\varphi_-=\frac{f(t-r/c)}{r},\qquad \varphi_+=\frac{f(t+r/c)}{r}.$$

If these functions are expanded into a power series in $1/c$, we obtain for the standing wave solution

$$\tfrac{1}{2}(\varphi_-+\varphi_+)=\frac{1}{r}\left[f(t)+\frac{1}{c^2}\frac{r^2}{2!}\frac{d^2}{dt^2}f(t)+\cdots\right].$$

Similarly, we obtain for the radiation terms

$$\tfrac{1}{2}(\varphi_--\varphi_+)=-\frac{1}{c}\left[\frac{d}{dt}f(t)+\frac{1}{c^2}\frac{r^2}{3!}\frac{d^3}{dt^3}f(t)+\cdots\right].$$

Thus, the starting term for the radiation terms is just the time derivative of the starting term for the standing wave solution. In general, however, the relationship is not as simple. Consider a function $g(r,t,\lambda)$ where $\lambda=v/c$. By generalizing a calculation by Page for the Liénard-Wiechert potentials[12] we find

$$g_\mp(r,t,\pm\lambda)=g(r,t,0)\pm\frac{1}{c}\left[\frac{dr}{dt}\left(\frac{d}{d\lambda}g(r,t,\lambda)\right)_{\lambda=0}-r\frac{d}{dt}g(r,t,0)\right]+\cdots,$$

where the subscripts $-$ and $+$ mean that the function is to be evaluated at the retarded and advanced times respectively. Clearly, in order to determine the relationship between the radiation terms and the standing waves, the explicit dependence of g on λ must be known. The EIH approximation method cannot reveal this

[12] L. Page, Phys. Rev. 24, 296 (1924).

dependence. Hence, it is not the appropriate formalism to use for an investigation of gravitational radiation and an approximation method which does not restrict the particle velocities must be found. The only purpose of the present paper is to show, contrary to previous results, that the EIH does not exclude the possibility of radiation.

In an accompanying paper,[13] Dr. Scheidegger remarks that our treatment of the transformation equations alters the physical situation because we neglect higher order solutions of the homogeneous field equations. On the other hand, keeping these terms violates the EIH method of approximation.[10] Therefore, he concludes that the starting function for the radiation terms must be rejected as a possible solution for freely interacting particles. It would appear, however, that a more reasonable solution of the problem, if the physical situation is altered by the transformations, would be to reject the transformations themselves. Indeed, Infeld[3] rejects a coordinate transformation which eliminates the perihelion precession of a double star system precisely because the transformation violates the EIH prescriptions. Since these prescriptions are rather stringent, it is hardly surprising that there exist some coordinate transformations which are not allowed. However, our purpose was to show that there exists a nonsingular solution for the radiation terms which cannot be removed by a coordinate transformation and which leads to the radiation of energy. Inasmuch as we know nothing about the form of the gravitational radiation terms, if any, the existence of a nonsingular solution which leads to radiative effects shows that radiation of energy by freely interacting masses cannot be ruled out. Only by finding another approximate solution for the gravitational field equations—one which does not depend on small particle velocities—can we determine whether or not freely interacting bodies radiate energy.

However, one can say that if gravitational radiation exists, it plays a very small role in the problem of motion. The EIH solution has been carried out to yield the equations of motion through the sixth order in v/c. This solution is sufficient to give the perihelion precession of a double star system.[14] On the basis of the present calculation, radiation effects should first appear in the eleventh order (one order above that in which the radiation of energy occurs). In addition Eq. (53) implies that, to be observable, the radiating system should not only have a large moment of inertia, but also a large rate of change for its moment of inertia. If these conditions are satisfied, the EIH approximation method may be poor and other methods will have to be found.

I should like to express my appreciation to Professor Bergmann for discussions on the significance of the EIH approximation method as well as for a critical reading of the manuscript.

[13] A. Scheidegger, following paper [Phys. Rev. **99**, 1883 (1955)].
[14] H. P. Robertson, Ann. Math **39**, 101 (1938).

APPENDIX 1. FIELD EQUATIONS

In all covariant field theories whose field equations are derivable from a variational principle, the strong energy-momentum psuedo-tensor has the form[15]

$$T_\mu{}^\nu=-F_{A_\mu{}^{B\nu}}y_B L^A+t_\mu{}^\nu. \tag{A1}$$

The $F_{A_\mu{}^{B\nu}}$ are certain constants, y_A are the field variables, L^A the field equations, and $t_\mu{}^\nu$ the weak energy-momentum psuedo-tensor defined by Eq. (4). Since the $T_\mu{}^\nu$ may be derived from the superpotentials,

$$T_\mu{}^\nu=U_\mu{}^{[\nu\sigma]}{}_{,\sigma}, \tag{A2}$$

it follows that linear combinations of the field equations may be expressed in terms of the superpotentials:

$$-F_{A_\mu{}^{B\nu}}y_B L^A=U_\mu{}^{[\nu\sigma]}{}_{,\sigma}-t_\mu{}^\nu. \tag{A3}$$

We shall find it convenient to use the field equations in this form. If, as in the general theory of relativity, the Lagrangian density is homogeneous quadratic in the first derivatives of the field variables,

$$L=\Lambda^{A\rho B\sigma}y_{A,\rho}y_{B,\sigma}, \tag{A4}$$

the weak energy-momentum psuedo-tensor takes the simple form:

$$t_\mu{}^\nu=-\delta_\mu{}^\nu\Lambda^{A\rho B\sigma}y_{A,\rho}y_{B,\sigma}+2\Lambda^{A\nu B\sigma}y_{A,\mu}y_{B,\sigma}. \tag{A5}$$

Specializing the above relations to the theory of gravitation, we find

$$-F_{(\alpha\beta)\mu}{}^{(\gamma\delta)\nu}g_{\gamma\delta}\sqrt{-g}G^{\alpha\beta}=2\sqrt{-g}G_\mu{}^\nu \tag{A6}$$

$$\Lambda^{(\alpha\beta)\rho(\gamma\delta)\sigma}=\frac{\sqrt{-g}}{8}\{g^{\alpha\beta}(g^{\gamma\rho}g^{\delta\sigma}+g^{\gamma\sigma}g^{\delta\rho})+g^{\gamma\delta}(g^{\alpha\rho}g^{\beta\sigma}+g^{\alpha\sigma}g^{\beta\rho})+g^{\rho\sigma}(g^{\alpha\gamma}g^{\beta\delta}+g^{\alpha\delta}g^{\beta\gamma}-2g^{\alpha\beta}g^{\gamma\delta})-[g^{\alpha\sigma}(g^{\beta\gamma}g^{\rho\delta}+g^{\beta\delta}g^{\rho\gamma})+g^{\beta\sigma}(g^{\alpha\gamma}g^{\rho\delta}+g^{\alpha\delta}g^{\rho\gamma})]\}, \tag{A7}$$

and[16]

$$U_\mu{}^{[\nu\sigma]}=\sqrt{-g}\{\delta_\mu{}^\nu(g^{\sigma\tau}g^{\rho\lambda}-g^{\lambda\sigma}g^{\rho\tau})-\delta_\mu{}^\sigma(g^{\nu\tau}g^{\rho\lambda}-g^{\nu\lambda}g^{\rho\tau})+\delta_\mu{}^\rho(g^{\nu\tau}g^{\lambda\sigma}-g^{\sigma\tau}g^{\lambda\nu})\}g_{\rho\lambda,\tau}. \tag{A8}$$

In terms of the $\gamma_\mu{}^\nu$, (A5) and (A8) become:

$$t_\mu{}^\nu=-\tfrac{1}{4}\delta_\mu{}^\nu(\gamma_{\alpha\beta,\rho}\gamma_{\alpha\beta,\rho}-2\gamma_{\alpha\beta,\rho}\gamma_{\rho\alpha,\beta}-\tfrac{1}{2}\gamma_{\tau,\rho}\gamma_{\tau,\rho})+\tfrac{1}{2}\eta^{\rho\nu}(\gamma_{\alpha\beta,\mu}\gamma_{\alpha\beta,\rho}-2\gamma_{\alpha\beta,\mu}\gamma_{\rho\alpha,\beta}-\tfrac{1}{2}\gamma_{\tau,\mu}\gamma_{\tau,\rho})+\text{terms of higher degree}, \tag{A9}$$

[15] P. G. Bergmann, Phys. Rev. **75**, 680 (1949).
[16] P. Freud, Ann. Math **40**, 417 (1939). The definition of $T_\mu{}^\nu$ given here differs from that of Freud by a factor of 2. Therefore, in reference 7 Eq. (2.23) should be multiplied by 2.

$$U_\mu{}^{[\nu\sigma]}=(\delta_\mu{}^\sigma\eta^{\nu\lambda}-\delta_\mu{}^\nu\eta^{\lambda\sigma})\gamma_{\rho\lambda,\rho}+(\eta^{\nu\tau}\eta^{\lambda\sigma}-\eta^{\nu\lambda}\eta^{\sigma\tau})\gamma_{\mu\lambda,\tau}$$
$$+(\delta_\mu{}^\nu\eta^{\sigma\tau}-\delta_\mu{}^\sigma\eta^{\nu\tau})(-\gamma_{\rho\lambda}\gamma_{\rho\lambda,\tau}+\gamma_{\tau\lambda}\gamma_{\rho\lambda,\rho}$$
$$-\tfrac{1}{2}\gamma\gamma_{\rho\tau,\rho}+\gamma_{\rho\lambda}\gamma_{\rho\tau,\lambda}+\tfrac{1}{2}\gamma\gamma_{,\tau}-\tfrac{1}{2}\gamma_{\rho\tau}\gamma_{,\rho})$$
$$+(\eta^{\sigma\lambda}\eta^{\nu\tau}-\eta^{\nu\lambda}\eta^{\sigma\tau})(\tfrac{1}{2}\gamma\gamma_{\rho\lambda,\tau}-\gamma_{\rho\tau}\gamma_{\rho\lambda,\rho}$$
$$-\gamma_{\lambda\rho}\gamma_{\rho\lambda,\tau}+\tfrac{1}{2}\gamma_{\rho\lambda}\gamma_{,\tau})$$
$$+\text{terms of higher degree.}\quad\text{(A10)}$$

In order to simplify the appearance of the above equations we have used the following notation:

$$\eta^{\rho\sigma}\gamma_{\mu\rho}\gamma_{\nu\sigma}=\gamma_{\mu\rho}\gamma_{\nu\rho}=\gamma_{\mu 5}\gamma_{\nu 4}-\gamma_{\mu 4}\gamma_{\nu 5},\quad\text{(A11)}$$

$$\gamma=\gamma_{\rho\rho}.\quad\text{(A12)}$$

The expansions in Eqs. (A9) and (A10) are sufficient to give us the field equations and surface integrals to all orders of interest. The field equations are obtained by substituting Eqs. (A6), (A9), and (A10) into (A3); the surface integrals, by substituting (A9) and (A10) into Eq. (3); and the radiation, by substituting (A9) into Eq. (20).

APPENDIX 2. CURVATURE TENSOR

In terms of the Christoffel symbols, the curvature tensor is

$$R_{\iota\kappa\lambda}{}^\nu=\left\{\begin{matrix}\nu\\\lambda\iota\end{matrix}\right\}_{,\kappa}-\left\{\begin{matrix}\nu\\\lambda\kappa\end{matrix}\right\}_{,\iota}-\left\{\begin{matrix}\nu\\\sigma\iota\end{matrix}\right\}\left\{\begin{matrix}\sigma\\\lambda\kappa\end{matrix}\right\}+\left\{\begin{matrix}\nu\\\sigma\kappa\end{matrix}\right\}\left\{\begin{matrix}\sigma\\\lambda\iota\end{matrix}\right\}.\quad\text{(A13)}$$

Introducing the $\gamma_{\mu\nu}$, we have

$$R_{\iota\kappa\lambda}{}^\nu=\tfrac{1}{2}\eta^{\nu\rho}[(\gamma_{\iota\rho,\lambda}-\gamma_{\iota\lambda,\rho}-\tfrac{1}{2}\eta_{\iota\rho}\gamma_{,\lambda}+\tfrac{1}{2}\eta_{\iota\lambda}\gamma_{,\rho})_{,\kappa}$$
$$-(\gamma_{\kappa\rho,\lambda}-\gamma_{\lambda\kappa,\rho}-\tfrac{1}{2}\eta_{\kappa\rho}\gamma_{,\lambda}+\tfrac{1}{2}\eta_{\lambda\kappa}\gamma_{,\rho})_{,\iota}]$$
$$+\text{terms of higher degree.}\quad\text{(A14)}$$

In particular, we find for the $(s00,r)$ component:

$$R_{s00}{}^r=-\tfrac{1}{2}(\gamma_{rs,00}-\gamma_{s0,0r}-\gamma_{r0,0s}-\tfrac{1}{2}\eta_{rs}\gamma_{,00}$$
$$-\tfrac{1}{2}\gamma_{,rs}+\gamma_{00,rs})+\cdots.\quad\text{(A15)}$$

APPENDIX 3. TABLE OF SURFACE INTEGRALS

The following table is helpful in evaluating the surface integral:

I. $$\frac{1}{4\pi r^2}\oint n_\sigma dS=0,$$

II. $$\frac{1}{4\pi r^2}\oint n_\tau n_\sigma dS=\frac{1}{3}\delta_{\tau\sigma},$$

III. $$\frac{1}{4\pi r^2}\oint n_\alpha n_\tau n_\sigma dS=0,$$

IV. $$\frac{1}{4\pi r^2}\oint n_\mu n_\alpha n_\tau n_\sigma dS=\frac{1}{15}(\delta_{\mu\alpha}\delta_{\tau\sigma}+\delta_{\mu\tau}\delta_{\alpha\sigma}+\delta_{\mu\sigma}\delta_{\alpha\tau}).$$

In the above, the n_σ are the components of the outward normal to the surface:

$$n_\sigma=x^\sigma/r$$

for a spherical surface.

APPENDIX 4. PROOF THAT ONE MAY CHOOSE $f_{\sigma\sigma}=0$

We start with the following solutions in the fifth and sixth orders:

$$_5\gamma_{rs}=f_{rs}{}^*(\tau),\quad\text{(A16a)}$$

$$_5\gamma_{00}=0,\quad\text{(A16b)}$$

$$_5\gamma_{0r}=0,\quad\text{(A16c)}$$

where $f_{rs}{}^*$ is a symmetric set of arbitrary functions of time. Consider the coordinate transformation generated by

$$_6v^0(\tau)=-\tfrac{1}{4}\int f_{nn}{}^*(\tau)d\tau,\quad\text{(A17a)}$$

$$_6v^r(\tau)=-\tfrac{1}{4}x^n f_{rn}{}^*(\tau).\quad\text{(A17b)}$$

From the transformation equations Eqs. (22), we find

$$_5\gamma'_{rs}=\tfrac{1}{2}f_{rs}{}^*-\tfrac{1}{2}\delta_{rs}f_{nn}{}^*\equiv f_{rs}(\tau),\quad\text{(A18a)}$$

$$_5\gamma'_{00}=0,\quad\text{(A18b)}$$

$$_5\gamma'_{0r}=\tfrac{1}{4}x^n f_{rn}{}^*{}_{,0}.\quad\text{(A18c)}$$

Clearly, $f_{nn}=0$; however, $f_{rs}{}^*$ still contributes to the solution through $_5\gamma'_{0r}$. The remainder of the discussion will show that solutions of the higher order field equations can be determined so that they depend only on f_{rs}.

In the seventh order the field equations are

$$_7\gamma_{rs,nn}-{}_7\gamma_{rn,sn}-{}_7\gamma_{sn,rn}+\delta_{rs}\,{}_7\gamma_{mn,mn}$$
$$=-{}_5\gamma_{0r,0s}-{}_5\gamma_{0s,0r}+2\delta_{rs}\,{}_5\gamma_{0n,0n}$$
$$+{}_5\gamma_{rs,00}-\delta_{rs}\,{}_5\gamma_{00,00}+\text{N.L.}\quad\text{(A19a)}$$

$$_7\gamma_{00,nn}={}_7\gamma_{mn,mn}+\text{N.L.}\quad\text{(A19b)}$$

with the coordinate conditions

$$_7\gamma_{rs,s}=0.\quad\text{(A19c)}$$

The nonlinear terms, N.L., contain products of $_5\gamma_{00}$ and $_5\gamma_{rs}$ only. Hence, they do not interest us because they depend on f_{rs} and not on $f_{rs}{}^*$. Therefore, after the substitution of Eqs. (18) the above equations become

$$_7\gamma'_{rs,nn}=f_{rs,00}{}^*+\text{N.L.},\quad\text{(A20a)}$$

$$_7\gamma'_{00,nn}=\text{N.L.}\quad\text{(A20b)}$$

The solutions of these equations consistent with the coordinate conditions [Eq. (A19c)]

$$_7\gamma'_{rs}=\frac{5}{22}r^2f_{rs,00}{}^*-\frac{1}{11}x^n(x^rf_{sn}{}^*+x^sf_{rn}{}^*)_{,00}$$
$$+\frac{3}{110}x^rx^sf_{nn}{}^*-\frac{1}{110}\delta_{rs}r^2f_{nn}{}^*+\text{N.L.},\quad\text{(A21a)}$$

$$_7\gamma'_{00}=\text{N.L.}\quad\text{(A21b)}$$

The following transformation in the sixth order removes ${}_6\gamma'_{0r}$:

$$ {}_6v^0=\tfrac{1}{8}x^m x^n f_{mn,0}{}^*. \tag{A22} $$

In the seventh order, we find [Eq. (22)]

$$ {}_7\gamma''_{rs}={}_7\gamma'_{rs}-\tfrac{1}{8}\delta_{rs}x^m x^n f_{mn,00}{}^*, \tag{A23a} $$

$$ {}_7\gamma''_{00}={}_7\gamma'_{00}-\tfrac{1}{8}x^m x^n f_{mn,00}{}^*, \tag{A23b} $$

and the coordinate conditions are

$$ {}_7\gamma''_{rs,s}=-\tfrac{1}{4}x^n f_{nr,00}{}^*. \tag{A23c} $$

The field equations in the eighth order are

$$ {}_8\gamma_{0r,nn}-{}_8\gamma_{0n,rn}={}_7\gamma_{rn,n0}-{}_7\gamma_{00,0r}+\text{N.L.}, \tag{A24} $$

with the coordinate condition

$$ {}_8\gamma_{0r,r}={}_7\gamma_{00,0}. \tag{A25} $$

From Eqs. (A23), we find for these equations

$$ {}_8\gamma''_{0r,nn}=-\tfrac{1}{4}x^n f_{rn,000}{}^*+\text{N.L.}, \tag{A26} $$

$$ {}_8\gamma''_{0r,r}=-\tfrac{1}{8}x^m x^n f_{mn,000}{}^*+\text{N.L.}, \tag{A27} $$

and the solutions

$$ {}_8\gamma''_{0r}=-\frac{1}{56}x^r x^m x^n f_{mn,000}{}^*-\frac{1}{56}x^n r^2 f_{nr,000}{}^* +\frac{1}{280}x^r r^2 f_{nn,000}{}^*+\text{N.L.} \tag{A28} $$

A final transformation eliminates the "linear" terms in ${}_7\gamma''_{00}$ and ${}_8\gamma''_{0r}$ and brings back the standard coordinate conditions in the seventh order:

$$ {}_7v^r=-\frac{1}{56}x^r x^m x^n f_{mn,00}{}^*-\frac{1}{56}x^n r^2 f_{nr,00}{}^* +\frac{1}{280}x^r r^2 f_{nn,00}{}^*. \tag{A29} $$

As a result of the above transformation, the solutions in the seventh and eighth orders are

$$ {}_7\gamma'''_{rs}=\frac{81}{308}r^2 f_{rs,00}{}^*-\frac{3}{154}x^n(x^s f_{rn}{}^*+x^n f_{sn}{}^*),_{00} +\frac{1}{77}x^r x^s f_{nn,00}{}^*-\frac{5}{308}\delta_{rs}r^2 f_{nn,00}{}^* -\frac{3}{14}\delta_{rs}x^m x^n f_{mn,00}{}^*+\text{N.L.}, \tag{A30a} $$

$$ {}_7\gamma'''_{00}=\text{N.L.}, \tag{A30b} $$

$$ {}_8\gamma'''_{0r}=\text{N.L.} \tag{A30c} $$

A little arithmetic shows that Eq. (A30a) may be written as

$$ {}_7\gamma'''_{rs}=\frac{27}{154}r^2 f_{rs,00}-\frac{1}{77}x^n(x^r f_{sn}+x^s f_{rn}),_{00} -\frac{1}{7}\delta_{rs}x^m x^n f_{mn,00}, \tag{A31} $$

where f_{rs} is defined by Eq. (A18a). Clearly, the solution no longer depends on $f_{rs}{}^*$.

APPENDIX 5. SOLUTION FOR THE TRANSVERSE-TRANSVERSE COMPONENTS

In this case the initial solutions in the fifth and sixth orders are

$$ {}_5\gamma_{rs}=f_{rs}(\tau), \tag{A32a} $$

$$ {}_5\gamma_{00}=0, \tag{A32b} $$

$$ {}_6\gamma_{0r}=0, \tag{A32c} $$

with (see Appendix 4)

$$ f_{ss}(\tau)=0. \tag{A33} $$

The coordinate transformation generated by the following functions eliminates ${}_5\gamma_{rs}$:

$$ {}_5v^r=\tfrac{1}{2}x^n f_{rn}(\tau). \tag{A34} $$

From the transformation equations, Eq. (22), we find

$$ {}_5\gamma'_{rs}=0, \tag{A35a} $$

$$ {}_5\gamma'_{00}=0, \tag{A35b} $$

$$ {}_6\gamma'_{0r}=-\tfrac{1}{2}x^n f_{rn,0}. \tag{A35c} $$

Comparing Eq. (A35c) with (A18c) we observe that the solutions differ only by a factor of two. Therefore, the remaining transformations required for this discussion will follow the pattern of the previous Appendix.

In the seventh order the field equations are [Eqs. (A19) and (A35)]

$$ {}_7\gamma'_{rs,nn}=f_{rs,00}, \tag{A36a} $$

$$ {}_7\gamma'_{00,nn}=0, \tag{A36b} $$

and the coordinate conditions are

$$ {}_7\gamma'_{rs,s}=0. \tag{A37} $$

If no additional harmonic functions are added to our solution, we obtain in a straightforward manner

$$ {}_7\gamma'_{rs}=\frac{5}{22}r^2 f_{rs,00}-\frac{1}{11}x^n(x^2 f_{ns}+x^s f_{nr}),_{00}, \tag{A38a} $$

$$ {}_7\gamma'_{00}=0. \tag{A38b} $$

This solution differs from that of Eq. (A31) only by a harmonic function.

With the choice of transformation function

$$ {}_6v^0=\tfrac{1}{4}x^m x^n f_{mn,0}, \tag{A39} $$

the longitudinal component ${}_6\gamma'_{0r}$ can be eliminated. In addition, in the seventh order we find

$$ {}_7\gamma''_{rs}={}_7\gamma'_{rs}+\tfrac{1}{4}\delta_{rs}x^m x^n f_{mn,00}, \tag{A40a} $$

$$ {}_7\gamma''_{00}=\tfrac{1}{4}x^m x^n f_{mn,00}, \tag{A40b} $$

and the coordinate conditions

$$ {}_7\gamma''_{rs}=\tfrac{1}{2}x^n f_{nr,00}. \tag{A40c} $$

An examination of the transformation equations, Eq. (22), shows that these solutions cannot be removed by any coordinate transformation in the seventh order.

From Eq. (A9) it follows that the solutions in the seventh order cannot give rise to radiation effects until the tenth order. Therefore, we require the solution in the eighth order. The field equations in the eighth order are [Eqs. (A24) and (A40)]

$$ {}_8\gamma''_{0r,nn}=\tfrac{1}{2}x^n f_{nr,000}, \tag{A41a} $$

and the coordinate condition is

$$ {}_8\gamma''_{0r,r}={}_7\gamma''_{00,0}=\tfrac{1}{4}x^m x^n \dot{f}_{mn,000}. \tag{A41b} $$

The solution of the above equations is easily found to be

$$ {}_8\gamma''_{0r}=(1/28)x^n x^m(x^m f_{nr}+x^r f_{nm}),_{000}. \tag{A42} $$

It is now possible to carry out another coordinate transformation which not only will eliminate ${}_7\gamma''_{00}$ and ${}_8\gamma''_{0r}$, but also returns the standard coordinate conditions in the seventh order. This transformation is characterized by

$$ {}_7v^r=(1/28)x^n x^m(x^m f_{nr}+x^r f_{nm}),_{00}. \tag{A43} $$

By means of Eq. (22), we find

$$ {}_7\gamma'''_{rs}=(12/77)r^2 f_{rs,00}-(18/77)x^n(x^r f_{ns}+x^s f_{nr}),_{00} +(3/7)\delta_{rs}x^m x^n f_{mn,00}, \tag{A44a} $$

$$ {}_7\gamma'''_{00}=0, \tag{A44b} $$

$$ {}_8\gamma'''_{0r}=0, \tag{A44c} $$

and the coordinate condition

$$ {}_7\gamma'''_{rs,s}=0. \tag{A45} $$

PHYSICAL REVIEW VOLUME 99, NUMBER 6 SEPTEMBER 15, 1955

Gravitational Radiation

ADRIAN E. SCHEIDEGGER
Dominion Observatory, Ottawa, Ontario, Canada

(Received April 19, 1955)

Recently, several solutions of the field equations of general relativity theory have been published which are based upon the Einstein-Infeld-Hoffmann approximation method and claim to represent radiative effects in the motion of freely gravitating particles. These solutions are analyzed and it is shown that none produced to date represents a satisfactory discussion of the possibility of radiation by freely gravitating particles.

The problem of the possibility of gravitational radiation by free particles is then investigated upon general grounds. It is shown that the possibility of such radiation depends on the definition of free particles. This definition depends on the (unproven) assumption that the Einstein-Infeld-Hoffmann method is, in its physical outcome, independent of the coordinate system. If this assumption were not true, this would constitute a severe limitation of the method and imply that all results have to be expressed in a standard coordinate system. Under these circumstances, the definition of free particles is obviously equivalent to a postulate of absence of radiation.

IT has been claimed by Infeld and the author[1-4] that gravitational radiation within the framework of the Einstein-Infeld-Hoffmann (EIH) formalism[5] of solving the field equations of general relativity theory does not correspond to radiation from a purely mechanical system. In fact, it was shown that the introduction of a "radiation" term (for the latter's definition see Scheidegger[3]) at any stage of the EIH procedure corresponds to altering the coordinate system at that stage of the procedure, and therefore it was reasoned that the solution could always be continued in such a manner so as not to contain any radiation at all. The last train of thought is dependent upon the assumption that the EIH procedure, including the prescription for the characterization of freely gravitating particles, is a consistent formalism, yielding equations of motion that are physically independent of the formal choice of the coordinate system. As the latter assumption has never been proven, the argument of Infeld and the writer suffers from the same deficiency.

Every once in a while some calculations are published which seem to show explicitly that gravitational

[1] L. Infeld and A. E. Scheidegger, Can. J. Math. **3**, 195 (1951).
[2] A. E. Scheidegger, Phys. Rev. **82**, 833 (1951).
[3] A. E. Scheidegger, Revs. Modern Phys. **25**, 451 (1953).
[4] L. Infeld, Can. J. Math. **5**, 17 (1953).
[5] See, e.g. A. Einstein and L. Infeld, Can. J. Math. **1**, 209 (1949).

radiation *does* exist in the framework of the EIH procedure. The "proof" is usually obtained by introducing a "radiation term" at some stage of the approximation, and by showing by a continuation of the approximation that the equations of motion are altered at some later stage in such a manner as to include radiative effects. The most important papers in this connection are by Hu[6] and by Goldberg.[7]

Although the matter has been discussed in some detail in earlier publications, the continued appearance of papers "proving" the existence of radiation seems to warrant a renewed discussion of those points in the author's earlier work that appear to be continually misunderstood.

Thus, let it be pointed out once more that it was never the author's intention to question the existence of solutions of the field equations of general relativity theory which represent radiative effects in the motion of the sources. The question is not whether gravitational radiation does exist or not; it is known that it may exist. The question is as to the physical meaning of such radiative effects, and thus as to whether two masses, left entirely alone after an arbitrary set of boundary conditions have been applied, can possibly radiate energy into infinity. In other words, the question is whether it is possible to choose such initial conditions that the masses, left alone thereafter, will radiate gravitational waves.

As the existence of gravitational radiation as such has never been questioned, the fact that authors like Hu and Goldberg were able to produce solutions that do represent such radiation, was to be expected. However, neither Hu nor Goldberg even touch upon a discussion of the physical conditions which those solutions represent; all they show is that there exist solutions of Einstein's field equations in which the particles follow a certain motion and, at the same time, energy is being radiated into infinity.

In fact, the problem of fixing the physical conditions prevalent in a certain solution of the field equations of general relativity, also occurs in the usual solution ordinarily obtained by the EIH method. At every stage of that procedure there is a considerable freedom of choosing solutions,—each represents a possible motion and a gravitational field belonging thereto. Therefore, prescriptions had to be found which would separate the solutions representing various physical conditions. In this instance, Einstein and Infeld[5] have given a set of prescriptions characterizing the "free" motion of two masses left alone, amongst which the most important one is that prohibiting the arbitrary choice of singularities, i.e., the arbitrary choice of solutions of the homogeneous field equations. The independence of these prescriptions of the coordinate system has never been proven, but one has generally assumed that they must be the same ones for every coordinate system, lest an unexpected limitation be inherent in the EIH method. If these prescriptions were not independent of the coordinate system (which is admittedly a possibility), then for a comparison of the physical significance of solutions, one would have to make coordinate transformations to identical systems.

In the consideration of any solutions of Einstein's field equations evolving from the EIH method, the foregoing remarks must be kept in mind.

Thus, let us suppose that the solutions published by Hu[6] and Goldberg[7] are mathematically correct, a supposition which the writer has not undertaken to investigate. One is then faced with the situation that one has two conflicting statements about the problem:

(i) a "proof" by Infeld and co-workers[1–4] that pairs of freely gravitating particles do not emit gravitational radiation. As stated above, this proof is tied up with a definition of "freely" gravitating particles which is based on the assumption that such a definition is independent of the choice of the coordinate system. For this assumption there has been no proof.

(ii) various solutions of the field equations which *do* give a radiation effect in the equations of motion.

These two statements about the problem constitute a dilemma. There are obviously two and only two ways out of this dilemma, viz., *Way(A)* by saying that the basis for the calculations of the radiation solutions is fallacious, *viz.*, by rejecting these particular solutions; *Way(B)* by saying that the basis for the Infeld argument is fallacious, *viz.*, by implying that the EIH formalism cannot be performed *ab initio* in any coordinate system such as to lead to the same physical conclusion.

It seems to the writer that one can accept "Way(B)" as a possibility out of the dilema only if "Way(A)" has been definitely ruled out.

Thus, let us first investigate *Way(A)* out of the dilemma. This involves an investigation of whether there are any reasons for which the radiation solutions of Hu[6] and Goldberg[7] would have to be rejected. In this instance, it must be noted that it cannot be considered as sufficient to make certain mathematical operations (such as Goldberg's[7] introduction of a radiation term plus coordinate transformations) without discussing whether this does not mean an alteration of the physical picture of free particles which has been postulated as basic in the Infeld argument. Since neither Hu[6] nor Goldberg[7] have even touched upon this question, neither of these authors has given an adequate substantiation of the claim that their solutions are admissible within the range of the Infeld definition of free particles. The writer feels that it is to the onus of any new investigator to substantiate his claims properly, rather than up to him to seek faults. However, it occurs to him that the discarding of some higher order terms by Goldberg in his coordinate transforma-

[6] N. Hu, Proc. Roy. Irish Acad. **A51**, 87 (1947).
[7] J. N. Goldberg, this issue [Phys. Rev. **99**, 1873 (1955)].

tions (reference 7, Sec. IV) may not be permissible although only (singular) solutions of the homogeneous equations are lost. It is precisely the solution of the homogeneous equation which determines the physical condition of free particles, although the latter does not change the equations of motion as calculated by Goldberg. Goldberg states that keeping the higher order terms (which *do* occur in the algorithm of the transformation) would violate the prescriptions of EIH. We agree with this and conclude that the coordinate transformation as handled by Goldberg in his deduction of radiation effects is probably in need of some further substantiation. It thus appears that "Way(A)" out of the dilemma has not been properly ruled out.

Next, let us examine "*Way(B)*" out of the dilemma. This implies an investigation of whether the Infeld argument is based upon too narrow assumptions. In order to do this, it is not sufficient to show that by keeping some higher order terms in a coordinate transformation one gets something different from what one gets by solving the equations anew in the new system after discarding the higher order terms,—even if this were done properly. Clearly, such a procedure just poses the problem, but certainly does not solve it. If the argument of Goldberg is right, there must be some intrinsic restriction in the EIH procedure which would be the real reason for the discrepancy. The argument of Goldberg (reference 7, Sec. IV) that the prohibition of solutions of the homogeneous equation is not a covariant requirement, would provide for such a restriction, but it is tied up with the admitting-procedure for the solutions. It still appears that if *all* solutions assumed as inadmissible according to "Way(A)" out of the dilemma are rejected, the requirements postulated for "free particles" are unique.

It thus appears that the two ways out of the dilemma actually cannot be considered separately. In other words, Infeld and co-workers considered a certain class of particle-motions, termed as "free," of which it was shown that they do not represent radiation. Therefore, anyone may insist that radiation does exist; a discrepancy exists only if it can be shown that such radiation exists among the class of motions considered by Infeld. Heretofore, nobody appears to have claimed this. The definition of this class of motions by Infeld presumes that it can be defined convariantly. This presumption, in turn, is based on the assumption that the EIH procedure is, in its physical outcome, independent of the chosen coordinate system. The latter assumption has never been proven, and it is thus possible that it might not be true. In that case, however, the definition of free particles would still be unique and consistent, but it would imply that one has to transform to a certain coordinate system for the investigation of the physical conditions represented by the mathematical solution under consideration. In this particular coordinate system, it is evident that the class of free motions is *eo ipso* equivalent to the class of nonradiative solutions.

Letters to the Editor

PUBLICATION of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is five weeks prior to the date of issue. No proof will be sent to the authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not exceed 600 words in length and should be submitted in duplicate.

Microwave Absorption in Compressed Oxygen

A. A. MARYOTT, *National Bureau of Standards, Washington, D.C.*

AND

G. BIRNBAUM, *National Bureau of Standards Boulder Laboratories, Boulder, Colorado*

(Received July 25, 1955)

THE microwave absorption in O_2, which is the result of the presence of a permanent magnetic dipole, consists of two parts.[1] The resonant contribution, governed by the selection rules $\Delta J=\pm 1$, $\Delta K=0$ where J is the quantum number for the total angular momentum and K is that for end-over-end rotation, contains a number of lines which, with one exception, lie in the vicinity of 2 cm^{-1}. Although this absorption has been investigated by a number of workers, the nonresonant, or Debye type, contribution corresponding to $\Delta J=0$, $\Delta K=0$ has not been the subject of prior investigation. With the aid of the Van Vleck-Weisskopf relation[2] the complete absorption coefficient α(cm^{-1}) is given by

$$\frac{\alpha}{2\pi\nu}=\frac{2\pi N\mu^2}{3kT}\left[\frac{2}{3}\left(\frac{\nu\Delta\nu_d}{\Delta\nu_d{}^2+\nu^2}\right)+\sum_{K\pm}\frac{f_{K\pm}|\mu_{K\pm}|^2}{\mu^2}\times\left(\frac{\nu\Delta\nu_{K\pm}}{\Delta\nu_{K\pm}{}^2+(\nu-\nu_{K\pm})^2}+\frac{\nu\Delta\nu_{K\pm}}{\Delta\nu_{K\pm}{}^2+(\nu+\nu_{K\pm})^2}\right)\right],$$

where $\mu=0.0262\times10^{-18}$ esu, the subscripts $K_\pm$ and d refer to the resonant and nonresonant terms, respectively, and the remaining symbols have their usual significance.

Measurements of the absorption were made at 25°C using high-Q resonant cavities at pressures up to 8 atmos for $\nu=0.0765$ cm^{-1}, and at pressures up to 40 atmos for $\nu=0.302$ and 0.778 cm^{-1}.

At the lowest frequency, where overlap from the resonant contribution is nearly negligible, the two parameters of the nonresonant term were evaluated. The value of $\tan\delta/p$ (maximum) is $2.9_1\times10^{-7}$ atmos^{-1} and in close agreement with the theoretical value of 2.87×10^{-7} atmos^{-1}. The line breadth parameter $\Delta\nu_d/p$ is 0.017_8 cm^{-1} atmos^{-1} and corresponds to an effective collision diameter of 2.6 A. This value is decidedly smaller than the average value of 4.4 A effective for resonant absorption[3] and also smaller than the kinetic collision diameter of 3.5 A.

Data obtained at 0.778 cm^{-1} are shown by curve 1 of Fig. 1. At this frequency substantial contributions arise from both the resonant and the nonresonant terms. Subtraction of the nonresonant part using the above-mentioned parameters gives curve 2. Above 10 atmos this curve lies above that computed for the resonant absorption (curve 3) by summing the contributions of the individual lines using recent data for the line breadth parameters.[3] At 40 atmos the value of $\tan\delta/p$ is more than twice that calculated, a discrepancy that is not materially reduced no matter what values are assumed for the resonant and nonresonant line widths. The nature of the discrepancy appears to be analogous to that previously observed for NH_3[4] and ND_3.[5] The data can be reconciled with the Van Vleck-Weisskopf line shape only if the resonant frequencies and line breadth parameters are both allowed to decrease with increasing pressure. For example, data at 0.302 cm^{-1} (not shown) and 0.778 cm^{-1} require average values of $\Delta\nu_{K\pm}/p$ and $\nu_{K\pm}$ of about 0.025 cm^{-1} atmos^{-1} and 0.5 cm^{-1}, respectively, in contrast to the initial values of 0.05 cm^{-1} atmos^{-1} and 2.0 cm^{-1}. A detailed report of this work will be published subsequently.

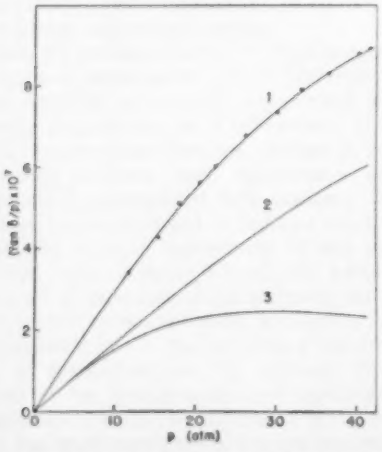


FIG. 1. Data for O_2 at $\nu=0.778$ cm^{-1} showing the experimental values of $\tan\delta/p$ (1) and the contribution associated with resonant absorption as derived from the experimental data (2) and from theory (3).

[1] J. H. Van Vleck, Phys. Rev. **71**, 413 (1947).
[2] J. H. Van Vleck and V. F. Weisskopf, Revs. Modern Phys. **17**, 227 (1945).
[3] J. O. Artman and J. P. Gordon, Phys. Rev. **96**, 1237 (1954).
[4] B. Bleaney and J. H. N. Loubser, Proc. Phys. Soc. (London) **A63**, 483 (1950).
[5] G. Birnbaum and A. A. Maryott, Phys. Rev. **92**, 270 (1953).

Infrared Absorption of Liquid and Solid Hydrogen with Various Ortho-Para Ratios*

W. F. J. Hare,† Elizabeth J. Allin, and H. L. Welsh
McLennan Laboratory, University of Toronto, Toronto, Canada
(Received July 21, 1955)

THE fundamental vibrational absorption of liquid and solid normal hydrogen has recently been studied in the temperature range, 10–22°K.[1] These experiments, designed to show the transition from collision-induced absorption in the high-pressure gas[2] to the corresponding phenomena in the condensed phases, have now been extended to various mixtures of ortho- and parahydrogen. Parahydrogen, prepared by the method described by Squires and Stewart,[3] was mixed with normal hydrogen at room temperature and the mixture condensed in the absorption cell previously described.[1]

Absorption profiles for solid hydrogen, 8.4 mm thick, with different ortho-para ratios are shown in Fig. 1. Three groups of main maxima are apparent: the Q group (Q_Q,Q_R) near the band origin, the $S(0)$ group [$S_1(0)$,$S_2(0)$,$S'(0)$] corresponding to the *para*-transition $J=0\rightarrow J=2$, and the $S(1)$ group [$S_1(1)$,$S_2(1)$, $S'(1)$] corresponding to the *ortho*-transition $J=1\rightarrow J=3$. The weak component, $S_2(0)$, is clearly resolved only in pure parahydrogen. The maxima designated as $2S(0)$, $S(0)+S(1)$, and $2S(1)$ are due to double rotational transitions; in parahydrogen only $2S(0)$ remains.

The broad components Q_R, $S'(0)$ and $S'(1)$ are clearly summation tones of the molecular vibrational and rotational frequencies with the vibrational frequencies of the crystal lattice. The shapes of these components are independent of, and their intensities vary linearly with the ortho-para ratio; thus, the transition probabilities (intensity/ortho or para concentration) are independent of the ortho-para ratio. The Q_R component can be separated into para and ortho contributions, and the transition probabilities are in the proportion $Q_R(\text{para}):Q_R(\text{ortho}):S'(0):S'(1)=10:19:22:11$. The Q_R component has its analog in the spectrum of the high pressure gas where it is interpreted as a "summation tone" of the vibrational frequency of the absorbing molecule with the relative kinetic energy of the colliding molecule and of the surrounding molecules; for this type of absorption close collisions, in the region of overlap forces, appear to be necessary.[2]

The intensities of the strong sharp lines Q_Q, $S_1(0)$ and $S_1(1)$ vary in a complex manner with the ortho-para ratio. In parahydrogen the Q_Q component has zero intensity, showing that the transition $J=0\rightarrow J=0$ is forbidden. When the intensities are corrected for overlapping with other components, and when the data for the Q_Q and $S(1)$ components are extrapolated, the transition probabilities for parahydrogen are approximately in the proportion $Q_Q:S_1(0):S_1(1)=6:5:9$. In the theory of van Kranendonk and Bird the transition probabilities for rotation-vibrational transitions induced by quadrupole interaction are given by $\times 3(J+1)(J+2)/2(2J+3)$ for $\Delta J=+2$ and $J(J+1)(2J+1)/(2J-1)(2J+3)$ for $\Delta J=0$. Since the observations in parahydrogen are in accordance with these it is concluded that the Q_Q, $S_1(0)$ and $S_1(1)$ components arise from quadrupole interaction. However, when the ortho-para ratio is not zero the observed relative transition probabilities differ widely from those given by the theoretical expressions. When the probabilities are extrapolated for the case of orthohydrogen they are in the proportion $Q_Q:S_1(0):S_1(1)=6:32:9$. The quadrupole effect thus seems to depend on whether the molecules in the neighborhood of the absorbing molecule are rotating or nonrotating.

The weak sharp lines $S_2(0)$ and $S_2(1)$ have the same frequency separation as the corresponding lines in the Raman effect of the gas[4]; for these, unlike the other components of the band, the B value is unperturbed. These lines may be due to the average quadrupole effect in the solid. The much stronger quadrupole effect giving rise to the Q_Q, $S_1(0)$ and $S_1(1)$ lines may perhaps be construed as a special two-body "encounter"

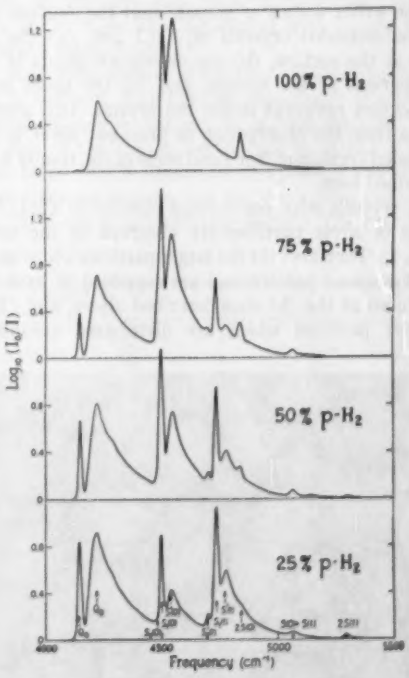


FIG. 1. The variation with ortho-para ratio of the fundamental infrared absorption of solid hydrogen at a temperature of 13.6±0.2°K.

of the absorbing molecule with one of its nearest neighbors, distinguished by the relative orientation of the quadrupole moments of the two molecules. These conclusions must be regarded as tentative.

* This research was supported by a grant from the National Research Council of Canada.
† Holder of a Research Fellowship of the National Research Council of Canada.

[1] Allin, Hare, and MacDonald, Phys. Rev. **98**, 554 (1955).
[2] D. A. Chisholm and H. L. Welsh, Can. J. Phys. **32**, 391 (1954).
[3] G. L. Squires and A. T. Stewart, J. Chem. Phys. **22**, 754 (1954).
[4] C. Cumming, Ph.D. thesis, University of Toronto, 1952 (unpublished).

Electron Traps in Silver Chloride

HIROSHI KANZAKI
Institute of Industrial Science, University of Tokyo, Chiba, Japan
(Received June 22, 1955)

EXPERIMENTAL evidence[1] is now available for trapping of electrons at dislocations in AgCl crystals. Seitz[2] proposed that the traps are incipient vacancies with an effective charge of $+e/2$, and his prediction has received further support in the work of Hedges and Mitchell[3] who have found that the silver particles produced by light outline the polygonized substructure in AgBr crystals.

The writer wishes to propose that the electron traps in well-annealed crystals of AgCl are: (a) the kink sites at the surface, (b) the emergence points of edge dislocations at the surface, and (c) the nodes in the dislocation networks inside the crystal. This proposal stems from the observation of print-out silver in well-annealed crystals of AgCl and some of the results will be described here.

In crystals with the surface parallel to (111),[4] two kinds of silver particles are observed at the surface (Fig. 1). They are: (1) the larger particles which outline the hexagonal pattern and are supposed to have been produced at the (b) sites described above, and (2) the smaller particles which are distributed around the hexagonal pattern and probably have been produced at the (a) sites.

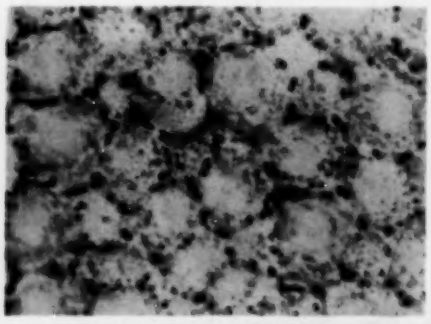


FIG. 1. Print-out silver particles at the surface of a AgCl crystal with (111) surface (×600).

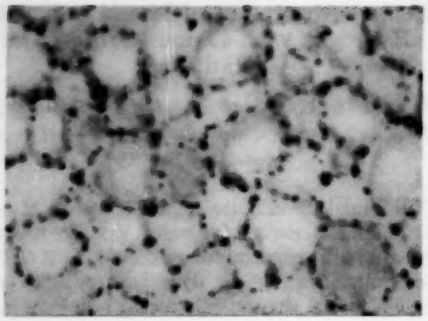


FIG. 2. Print-out silver particles inside the AgCl crystal with (111) surface (×600).

In the interior, the particles are distributed discontinuously upon the hexagonal cylindrical surfaces outlined by the particles of (1) at the surface (Fig. 2) and are supposed to have been produced at the nodes of dislocation networks extended on the cylindrical surfaces.

This selective production of silver particles may be explained by the trapping of electrons by effective charges. In the case of traps at the (b) sites, the extra (110) half-plane accompanied by the edge components of dislocations may produce extra charge by intersecting (111) and (100) surface planes, but cannot do this at the (110) surface plane. This prediction has been confirmed by the following experiments.

Figure 3 shows the distribution of silver particles near the grain boundary. The right-hand grain with (100) surface has been subjected to extensive darkening and the left-hand one with (110) surface has produced very little silver. Chemical etching[5] revealed etch pits due to dislocations (probably with edge components) in the (110) grain but not in the (100) grain. Figure 4 shows the etching pattern near the grain boundary of

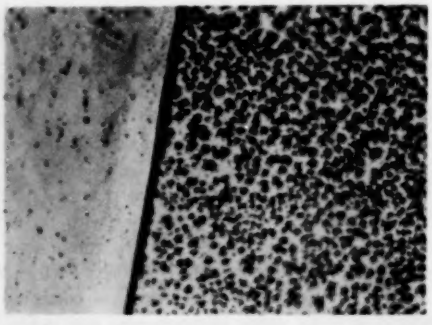


FIG. 3. Print-out silver particles near the grain boundary. The right-hand crystal has a (100) surface and the left-hand crystal has a (110) surface (×320).

FIG. 4. Chemical etching pattern near the same boundary as shown in Fig. 3 (×320).

Fig. 3. It may be concluded that the edge dislocations in the (110) grain have not contributed to the formation of silver particles.

The density of dislocations in well annealed AgCl has been found to be 10^6 or 10^7 per cm^2 in our experiments.

The writer wishes to express his gratitude to Professor Y. Tani for his valuable guidance and also to the cordial hospitality of the Research Laboratories, Fuji Photo Film Company, Ltd., in the course of this work.

[1] J. R. Haynes and W. Shockley, Phys. Rev. **82**, 935 (1951); F. C. Brown, Phys. Rev. **97**, 355 (1955).
[2] F. Seitz, Revs. Modern Phys. **23**, 328 (1951).
[3] J. M. Hedges and J. W. Mitchell, Phil. Mag. **44**, 223, 357 (1953).
[4] The surface of the crystals grown from the melt was usually parallel to (100) or (110). The crystals with (111) surface were prepared under special conditions of crystal growth.
[5] Etching was carried out for a few minutes by the FF-H photographic fixer (Fuji Photo Film Company, Ltd., Kanagawa, Japan).

Thermoelectric Power of Indium Antimonide

H. P. R. FREDERIKSE AND EUGENIE V. MIELCZAREK
National Bureau of Standards, Washington, D. C.
(Received July 11, 1955)

THE thermoelectric power of single crystals of InSb has been measured on two samples between 60 and 400°K. The four-probe method used for the determination of this quantity has been described in recent years by several authors.[1-3] The Hall coefficient and resistivity of the specimen were also determined. One sample appeared to be p-type and had an effective impurity concentration of 6.7×10^{15} per cc; the hole mobility at 80°K was 3300 cm^2/volt-sec. The other sample showed electron conduction; its carrier concentration at low temperature (80°K) was 7×10^{15} per cc, and the electron mobility 10^5 cm^2/volt-sec.

Figure 1 shows the thermoelectric power Q as a function of temperature. In the intrinsic range this quantity can be described by the following formula[4]:

$$Q=-\frac{k}{e}\left[\frac{b-1}{b+1}\left(\frac{E_G}{2kT}+2\right)+\tfrac{3}{4}\ln\frac{m_1^*}{m_2^*}\right], \qquad (1)$$

where k = Boltzmann constant, b = mobility ratio, E_G = energy gap at T°K, and m_1^*, m_2^* = effective masses of electrons and holes respectively. Taking $b=35$ and E_G (at 300°K) = 0.16 ev,[5] we find $m_1^*/m_2^*=0.11$.

Calculations of the thermoelectric power in the impurity range, when lattice scattering is predominant, yields the following result[4]:

$$Q=\pm\frac{k}{e}\left[2-\ln\frac{n_{1,2}h^3}{2(2\pi m_{1,2}^*kT)^{\frac{3}{2}}}\right], \qquad (2)$$

where $n_{1,2}$ = number of free carriers (electrons or holes); the plus or minus sign refers to p- or n-type respectively. At low temperatures one has to consider the increased influence of impurity scattering which will change the expression (2) slightly. For reasons explained below, only the results for the n-type sample between 100 and 200°K can be compared with theoretical predictions. Calculation of the effective mass of the electron at temperatures between 160 and 200°K yields consistently the value $m_1^*=0.014m_e$. Combining this value with the above-mentioned result for the ratio m_1^*/m_2^*, one obtains an effective hole mass $m_2^*=0.13m_e$. These values are in reasonable agreement with results from microwave measurements.[6]

An interesting feature of the thermoelectric power of InSb is the fact that the "phonon-drag" effect[2,7] is observed in p-type but not in n-type samples. The steep rise of the thermoelectric power of the p-type sample below 150°K illustrates this point. (The rather high degeneracy temperature for the n-type sample—about 90°K—is a complicating factor.) The contribu-

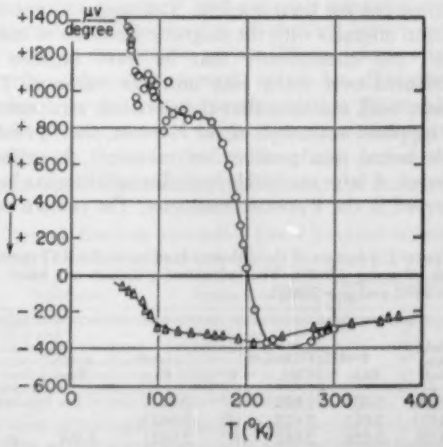


FIG. 1. Thermoelectric power of InSb. ○ p-type; △ n-type.

tion to the thermoelectric power due to electron-phonon interaction is proportional to $l_\phi'/l_{e\phi}$;[3] the numerator l_ϕ' represents the mean free path of those phonons which interact with electrons, while $l_{e\phi}$ is the free carrier mean free path determined by collisions with phonons. From the mobilities one calculates that $l_{e\phi}$ at 150°K is 5×10^{-4} cm for holes and 6×10^{-5} cm for electrons. The average mean free path of all the phonons can be evaluated from thermal conductivity data[6]; assuming that the mean free path for long wavelength phonons is about 10 times larger, one finds for l_ϕ' a value of 10^{-5} to 10^{-4} cm. Calculation then shows that the phonon-hole interaction yields already a measurable contribution to the thermoelectric power at 150°K, while that due to the phonon-electron interaction is quite negligible.

Note added in proof.—The values for the effective masses have been calculated taking the proper correction for impurity scattering into account.

[1] W. W. Scanlon, Phys. Rev. **92**, 219 (1953).
[2] W. W. Tyler and A. C. Wilson, Jr., Phys. Rev. **89**, 870 (1953).
[3] H. P. R. Frederikse, Phys. Rev. **92**, 248 (1953).
[4] V. A. Johnson and K. Lark-Horovitz, Phys. Rev. **92**, 226 (1953).
[5] Breckenridge, Blunt, Hosler, Frederikse, Becker, and Oshinsky, Phys. Rev. **96**, 571 (1954).
[6] R. N. Dexter and Benjamin Lax, Phys. Rev. **99**, 635(A) (1955).
[7] Conyers Herring, Phys. Rev. **96**, 1163 (1954).
[8] G. Busch and M. Schneider, Physica **20**, 1084 (1954).

Electron Spin Resonance of V_1-Centers*

WERNER KÄNZIG
Department of Physics, University of Illinois, Urbana, Illinois
(Received August 1, 1955)

THE electron spin resonance of the F-center (electron bound to a negative ion vacancy) has been investigated extensively in the past.[1] The hyperfine splitting has not been resolved. This suggests that the electron interacts with the magnetic moments of many nuclei and consequently that its wave function is distributed over many ions near the vacancy. The present work indicates that the electronic structure of the supposed antimorph of the F-center, the V_1-center (hole bound to a positive ion vacancy[2]), is entirely different. A large resolvable hyperfine splitting has been observed in the V_1-center resonance. The pattern permits a detailed analysis of the electronic structure of this center and we have concluded that the hole is localized on but two negative ions.

TABLE I. g-factors of the different families in the KCl spectra. Limit of error ±0.001. The calculated g-factors are based on $g_\perp = 2.042$ and $g_{||} = 2.0023$.

Field direction	$\theta=45°$, 35° and 60° Exp.	Calc.	θ	$\theta=90°$ Exp.	$\theta=0°$ Exp.	
(100)	2.024	2.022	45°	2.042	...	...
(111)	2.015	2.015	35°	2.042 $g_\perp$	...	...
(110)	2.032	2.032	60°	2.042	2.002	$g_{||}$

TABLE II. Average spacing in gauss between the groups in the different families of the KCl spectra. Limit of error ±0.5 gauss.

Field direction	$\theta=45°$, 35°, and 60°	θ	$\theta=90°$	$\theta=0°$
(100)	69.6	45°	10.5	...
(111)	83.0	35°	10.0	...
(110)	54.2	60°	10.0	102.0

KCl, NaCl, KBr, and LiF single crystals were placed in a resonant cavity and irradiated with x-rays at −180°C. Then, without warming the crystal, we investigated the magnetic resonance spectrum with a high sensitivity resonance spectrometer operating at about 9300 Mc/sec. A modulation of the dc magnetic field of 0.1 to 1.0 gauss was used. The crystals were then warmed to a selected temperature and cooled again to −180°C. Upon remeasurement of the spectrum it was found that all lines had faded in the same proportion. This indicates that every spectrum is essentially due to a single kind of trap. The temperature region in which fast fading occurs coincides in the case of KCl and KBr with the temperature region in which the V_1-band bleaches.[3] In NaCl, fast fading occurs between −140°C and −120°C, in LiF between −160°C and −140°C. No optical data are available for these two substances.

The resonance spectra depend very sensitively upon the orientation of the crystals in the magnetic field. The cases $H\|(100)$, $H\|(111)$ and $H\|(110)$ have been investigated. In the case of KCl the spectra can be decomposed into families of lines, each family consisting of seven almost equally spaced groups. The integrated intensity ratio of the groups is 1:2:3:4:3:2:1. The number of families is 2 for $H\|(100)$ and $H\|(111)$, and 3 for $H\|(110)$. The g-factors corresponding to the center lines of the families are larger than the free-electron value except in the case of one of the families in the $H\|(110)$ pattern, for which g is equal to the free-electron value within experimental error. As the nuclear spin of Cl is $\frac{3}{2}$, the number of the groups (7) and their intensity ratio indicates that the hole spends its time near two halogens only, i.e., the V_1-center is essentially a Cl_2^- molecule-ion which is presumably near a positive-ion vacancy. The observed splitting agrees satisfactorily with the theoretical splitting calculated under the assumption of dipolar interaction by using Sternheimer's[4] $\langle 1/r^3\rangle$ values of the halogen p-functions. The g-shifts as well as the splittings can be explained quantitatively if we assume axially symmetric Cl_2^- molecule-ions oriented in the (110) directions. The distribution over the six different (110) directions is statistical. Tables I and II show as an example the main resonance data for KCl.

TABLE III. Approximate line width in gauss.

	KCl					
	$Cl^{35}-Cl^{35}$	$Cl^{35}-Cl^{37}$	$Cl^{37}-Cl^{37}$	NaCl	KBr	LiF
Line width in gauss	2.4	2.0	1.6	7.5	4	17

The variation of the splitting with the angle θ between the bond direction and the magnetic field enables us to determine the admixture of s and p functions on the halogens. We have obtained 30% s and 70% p for both KCl and NaCl. For $|\psi(0)|^2$, we have obtained 1.9×10^{24} cm^{-3}.

The KBr spectra are more difficult to analyze because of second-order effects and quadrupole effects. However, the analysis indicates the existence of Br_2^- molecule-ions. Second-order effects play also an important part in LiF. A tentative analysis suggests that the bond direction of the F_2^- molecule-ion is (110).

Typical values of the observed line widths are listed in Table III. It appears that the predominant relaxation is due to hyperfine interaction. Spin-spin interaction seems also to play a role, and this indicates that high local concentrations of color centers might be formed by x-irradiation.

A full account of the present investigations will be published in a forthcoming paper. The analysis of the results was only possible through the advice of C. P. Slichter and the continuous help of T. G. Castner.

* Partially supported by the Office of Naval Research.

[1] See, e.g., Kip, Kittel, Levy, and Portis, Phys. Rev. **91**, 1066 (1953).
[2] F. Seitz, Revs. Modern Phys. **26**, 7 (1954).
[3] D. Dutton and R. Maurer, Phys. Rev. **90**, 126 (1953).
[4] R. Sternheimer, Phys. Rev. **84**, 244 (1953).

Effect of Proton Irradiation upon the Electrode Potential of Tungsten

M. SIMNAD AND R. SMOLUCHOWSKI
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received July 18, 1955)

NUCLEAR irradiation may be expected to alter the electrochemical and chemical properties of metals. As a preliminary study we have measured the effect of proton irradiation upon the electrode potential of tungsten in aqueous salt solutions. Tungsten wire specimens 0.012 in. in diameter were chosen since previous investigations[1] have shown that at room temperature the radiation effects in this material are quite stable. Reproducible and steady electrode potential measurements were obtained by means of the following procedure. The tungsten surfaces were degreased with alcohol and acetone, immersed in a strong aqueous solution of ammonia for several minutes, and washed with an oxygen-free saturated solution of KCl in distilled water. The specimens were then placed in an oxygen-free saturated solution of KCl in distilled water, and their electrode potentials were measured against a saturated calomel cell with a vacuum-tube potentiometer. The solution and the standard cell were kept at 30°C in a thermostat. The results obtained are shown in Table I.

First of all it should be pointed out that after bombardment the wires were very slightly radioactive (about 1000 counts per minute), presumably because of some minor impurities. This activity is much too weak to produce any significant changes in the electrolyte and thus the observed effects have to be ascribed to radiation-induced changes in the metal. It appears that the effect of proton irradiation upon the electrode potential of tungsten is large and actually it is much greater than the effect of severe cold work. At first glance the magnitude of the observed effect is surprising in view of the fact that the diameter of the wire is less than one percent of the mean free path of the 260-Mev proton in tungsten. It should be remembered, however, that the majority of the radiation effects produced under these conditions appears to be caused by highly effective secondary nucleons produced in inelastic collisions rather than by elastic collisions of the incident protons.[1] The results indicate also that the change in electrode potential increases with the intensity of the proton beam as expected. The greater change obtained with protons of 130-Mev energy than with 260-Mev protons is rather difficult to understand and may not be real. In all cases the potentials of the irradiated specimens are anodic (less noble) to the annealed specimens, which means that irradiation makes the tungsten more reactive chemically.

The experiments show that irradiation alters significantly the electrochemical properties of metals, and that such measurements may be used to study the nature and extent of the effects produced. One would expect the change of the electrode potential to be associated with production of localized lattice imperfections at the metallic surface. These defects are primarily dislocations formed either by displacement spikes or by collapse of vacancy clusters. Other defects such as vacancies and interstitials, even if they were sufficiently stable in the interior of the grains, would probably disappear rather rapidly near the surface. The

TABLE I. Electrode potentials of proton irradiated tungsten.

Treatment	Proton irradiation	Electrode potential[a] millivolts (anodic)
Annealed[b]	none	0
Cold drawn	none	22
Annealed and irradiated	1.9×10^{13}/cm^2 at 130 Mev	72?
Annealed and irradiated	1.8×10^{13}/cm^2 at 260 Mev	39
Annealed and irradiated	6.8×10^{13}/cm^2 at 260 Mev	47
Annealed and irradiated	2.2×10^{14}/cm^2 at 260 Mev	84
Irradiated and annealed[b]	2.2×10^{14}/cm^2 at 260 Mev	0

[a] The electrode potentials are reported *versus* the annealed tungsten.
[b] The annealed specimens were heated for 2 hours at 900°C in an argon atmosphere.

existence of such a gradient of defects near the surface is indicated by the recent observations of the effect of irradiation on the solubility rate of Fe_2O_3 in hydrochloric acid.[2]

The authors wish to acknowledge the assistance of Dr. E. Pearlstein, who irradiated the specimens at the Carnegie Institute of Technology synchrocyclotron and of Mrs. Aija Spilners in measuring the electrode potentials. Support of this work has been given by a research program sponsored by the U. S. Atomic Energy Commission.

[1] Pearlstein, Ingham, and Smoluchowski, Phys. Rev. 98, 1530 (1955).
[2] M. Simnad and R. Smoluchowski, J. Chem. Phys. (to be published).

Radiative Transitions in Semiconductors

R. Braunstein
RCA Laboratories, Princeton, New Jersey
(Received July 19, 1955)

RADIATION produced by carrier injection has been observed from GaSb, GaAs, InP, and the Ge-Si alloys at room temperature and 77°K. The spectral distributions of the radiation are maximum at energies close to the best estimates of the band gaps of these materials; consequently, the evidence is that the radiation is due to the direct recombination of electron-hole pairs. Direct electron-hole recombination radiation from a semiconductor has been previously observed from Ge and Si by Haynes and Briggs[1] and from SiC by Lehovec *et al.*[2]

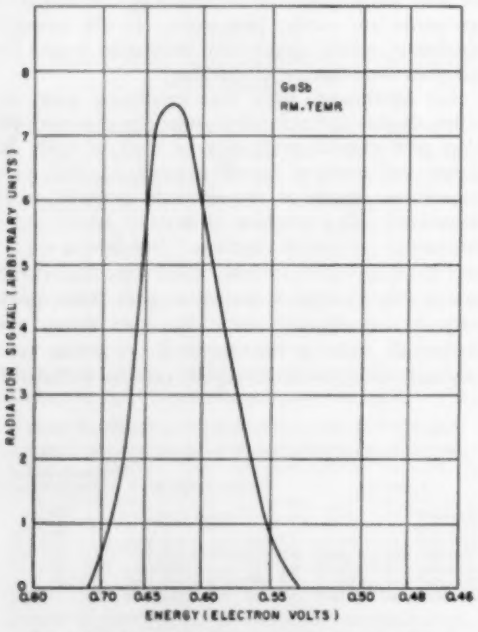


FIG. 1. Spectral distribution of recombination radiation from GaSb at room temperature.

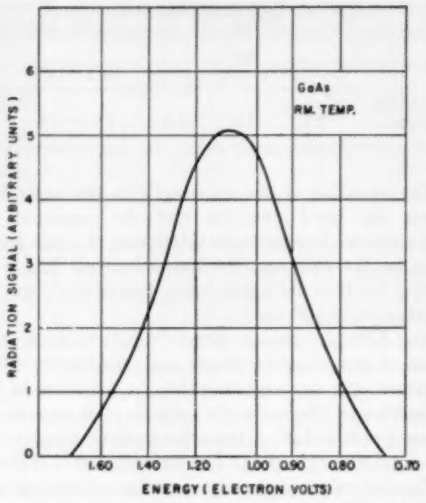


FIG. 2. Spectral distribution of recombination radiation from GaAs at room temperature.

The radiation was obtained by injecting minority carriers into 0.02 in. thick samples of material by means of point contacts or by broad area injecting contacts of silver paint applied to the surface of an appropriately etched semiconductor. The samples were soldered to kovar rings and mounted on temperature-controlled copper blocks. The diodes were pulsed by a 50% on-off square wave generator operating at 100 cps; currents up to 1 amp/cm² were passed in the forward direction. Radiation was detected by a dry-ice cooled PbS cell. The spectral distribution of the radiation from GaSb and GaAs was obtained by using a Perkin-Elmer monochromator with a fused quartz prism.

Figure 1 shows the spectral distribution of the radiation from GaSb at room temperature. The peak of the emission line occurs at 0.625 ev. This is to be compared with the value of 0.67 ev for the band gap of GaSb obtained by Blunt *et al.*[3] from absorption measurements. The present sample was *n*-type having a free carrier density of $\sim 10^{17}/cm^3$.

Figures 2 and 3 show the lines observed from GaAs at room temperature and 77°K, respectively. The peak of the radiation occurs at 1.10 ev at room temperature and at 1.19 ev at 77°K. If we assume a linear variation of the line position with temperature, the line position is given by $1.22-4\times10^{-4}T$ electron volt. If we attribute the radiation to the recombination of injected holes and electrons with the emission of a photon and the simultaneous absorption or the emission of a phonon, the line positions are at variance with the value of 1.35 ev

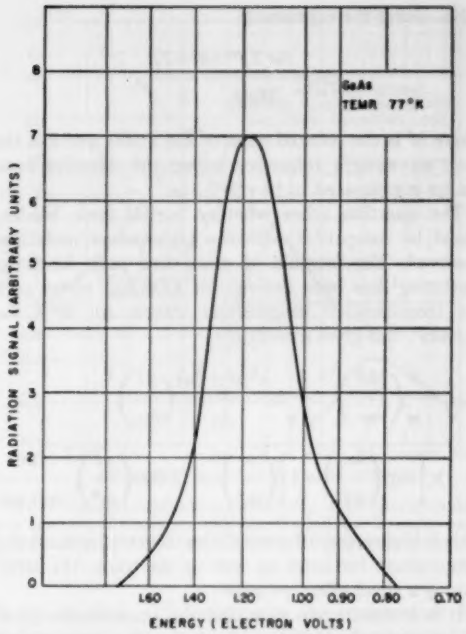


FIG. 3. Spectral distribution of recombination radiation from GaAs at 77°K. This line is to be compared with the line in Fig. 2 taken at room temperature. As the temperature is lowered, the peak intensity of the line increases, the width at half-maximum decreases, and the line shifts to higher energies.

for the room temperature band gap of GaAs as reported by Welker.[4] The absorption or the emission of a phonon can shift a line by only a few hundredths of an electron volt. Self-absorption of the radiation in the sample is also not sufficient to account for the apparent discrepancy. The sample of GaAs was *n*-type and had a free carrier density of $\sim 10^{17}/cm^3$.

The emission lines of InP and the Ge-Si alloys were not obtained. However, by appropriate filters, the radiation peaks were shown to occur at energies close to the band gaps of these materials.

The intensity of the radiation in all cases was observed to be a linear function of the injection current for the range of forward currents used (up to 1 amp/cm^2). The radiative recombination of electron-hole pairs is a bimolecular process and thus should be proportional to the product of the electron-hole concentrations. However, for the forward currents used, the majority carrier densities remained essentially constant and consequently to a first approximation the radiation will appear to very linearly with the minority carrier densities. In all cases, the integrated intensity of the radiation increased as the temperature of the sample was lowered.

Further results on radiative transitions in semiconductors will be presented at a later date.

The author wishes to express his thanks to Dr. D. A. Jenny for making available the samples of the intermetallic compounds and to Dr. E. G. Linder and Dr. A. R. Moore for stimulating discussions related to the problem.

[1] J. R. Haynes and H. B. Briggs, Phys. Rev. **86**, 647 (1952).
[2] Lehovec, Accordo, and Jamgochian, Phys. Rev. **83**, 603 (1951).
[3] Blunt, Hosler, and Frederikse, Phys. Rev. **96**, 576 (1954).
[4] H. Welker, Physica **20**, 893 (1954).

Infrared Absorption in Indium Antimonide*

W. G. SPITZER AND H. Y. FAN
Purdue University, Lafayette, Indiana
(Received July 25, 1955)

THE optical transmission of indium antimonide has been measured from 5 to 150 microns and at different temperatures. A residual ray monochromator was used for wavelengths beyond 35 microns. A number of lattice absorption bands were observed which indicate that there is some ionic binding in the crystal.

Figure 1 shows the transmission curves for a single crystal specimen with an acceptor concentration, N_A, of 10^{15} cm^{-3}. At 297°K, there are two absorption bands at 28.2 and 30 microns. With decreasing temperature these bands shift slightly to shorter wavelength. Measurements on samples of different impurity concentrations varying between 10^{15} to 10^{17} cm^{-3} show that the magnitudes of these bands are independent of carrier or impurity concentrations, indicating that these bands are inherent to the lattice.

Figure 2 shows the reflectivity curves at long wavelength for two temperatures. There is a strong band near 52 microns at both temperatures. While in the 78°K curve the reflectivity drops to 35% at longer wavelengths, the 297°K curve shows another rise of the reflectivity beyond 100 microns. These results show that

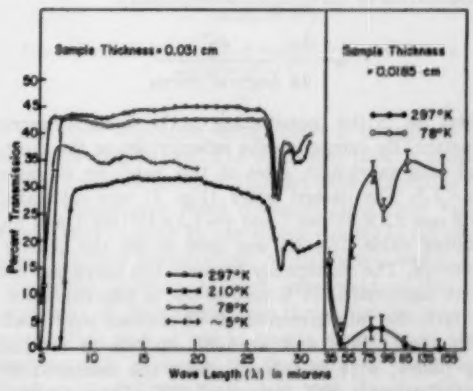


FIG. 1. Transmission as a function of wavelength for an InSb sample; $N_A = 10^{15}$ cm^{-3}.

there is a strong absorption band near 52 microns at both temperatures. Furthermore, there is also high absorption beyond 100 microns at 297°K. This is borne out by the transmission curves shown in Fig. 1. The sample is nearly intrinsic with $n=p=2.4\times10^{16}$ cm^{-3} at 297°K, while $p\sim10^{15}$ cm^{-3} and $n\ll10^{15}$ cm^{-3} at 78°K. The 52-micron band is thus independent of the carrier concentration and is therefore a lattice band. The high absorption beyond 100 microns, observed at 297°K seems to be caused by the carriers,[1] since the absorption becomes small at 78°K.

Reflectivity measurements at 297°K on an n-type sample of $n=10^{17}$ cm^{-3} and a p-type sample of $p=10^{17}$ cm^{-3} confirm the above interpretation. The same 52-micron band is present in both samples. Beyond 100 microns, the n-type sample shows an even higher reflectivity than the pure sample at 297°K while the p-type sample has the low reflectivity shown by the pure sample at 78°K. Thus we see that the high reflectivity at long wavelengths is caused by absorption due to the conduction electrons.

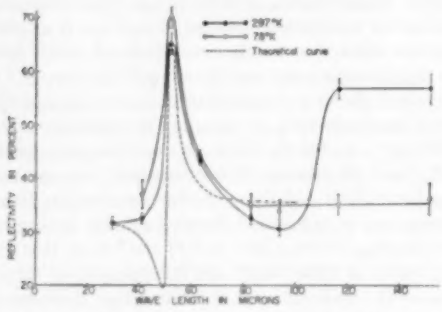


FIG. 2. Reflectivity as a function of wavelength for an InSb sample; $N_A=10^{15}$ cm^{-3}.

The strong lattice band near 52 microns indicates some degree of ionic binding. This absorption gives a contribution to the complex susceptibility,

$$\alpha=\frac{\Delta\epsilon_s}{4\pi}\frac{\omega_0^2}{(\omega_0^2-\omega^2)+i\gamma\omega},$$

where $\Delta\epsilon_s$ is the contribution to the static dielectric constant. By comparing the reflectivities on the short- and long-wavelength sides of this band we estimate $\Delta\epsilon_s<1.5$. The dashed curve (Fig. 2) was calculated with $\omega_0=3.5\times10^{13}$ sec^{-1} and $\gamma=1.3\times10^{12}$ sec^{-1}, and the limiting value $\Delta\epsilon_s=1.5$ was used to get the high reflectivity. The discrepancy between this curve and the curve measured at 78°K may be due to poor resolution. In fact, the measurements at 52 microns were made with three NaCl and an InSb crystals as residual ray plates; with four NaCl plates the measured reflectivity was ~50% instead of 65%. These observations confirm the narrowness of the reflection band in InSb. Using the expression[2]

$$\Delta\epsilon_s=\frac{4\pi Ne^{*2}}{M\omega_0^2}\frac{(n^2+2)^2}{3},$$

where M is the reduced mass of the atoms and n is the short-wavelength refractive index, the effective ionic charge is estimated to be $e^*=0.34e$.

The question arises whether partial ionic binding would be compatible with the high carrier mobilities observed. The formula of mean free path for polar scattering[3] has been derived for $kT\ll\hbar\omega_l$, where ω_l is the frequency of longitudinal waves. At 80°K, $\hbar\omega_l=3.4kT$, this gives a mobility

$$\mu_{\text{polar}}=\frac{e}{m}\left(\frac{3kT}{m}\right)^{-\frac{1}{2}}\frac{6}{\sqrt{\pi}}\frac{n^2(n^2+\Delta\epsilon)}{a_0\Delta\epsilon}\left(\frac{kT}{\hbar\omega_l}\right)^{\frac{1}{2}}$$
$$\times\left[\exp\left(\frac{\hbar\omega_l}{kT}\right)-1\right]\left(\frac{m}{m^*}\right)^{\frac{1}{2}}\simeq15\,000\left(\frac{m}{m^*}\right)^{\frac{1}{2}}\frac{\text{cm}^2}{\text{volt sec}}.$$

This is higher than the mobilities observed around this temperature for holes as well as electrons, the latter having a small effective mass.

It is interesting to note that the transmission up to 35 microns decreased with decreasing temperature within the extrinsic range; at 297°K the sample is intrinsic. As reported previously,[4] in samples with $N_A\gtrsim10^{17}$ cm^{-3} the transmission decreases steadily from 297°K down.

This work was stimulated by some preliminary reflection measurements made by E. J. Johnson at the suggestion of Professor K. Lark-Horovitz.

* Work supported by a Signal Corps Contract.

[1] H. Y. Fan and M. Becker, *Symposium Volume of the Reading Conference* (Butterworth Publishing Company, London, 1951).
[2] H. Fröhlich, *Theory of Dielectrics* (Clarendon Press, Oxford, 1949).
[3] Fröhlich, Pelzer, and Zienau, Phil. Mag. **41**, 211 (1950).
[4] W. Kaiser and H. Y. Fan, Phys. Rev. **98**, 966 (1955).

Experimental Evidence for Dislocations in Crystalline Quartz

H. E. BÖMMEL, W. P. MASON, AND A. W. WARNER, JR.
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received July 26, 1955)

RECENT measurements of well etched, mounted, and contoured AT shear vibrating quartz crystals, shown by Fig. 1, indicate that the Q decreases inversely proportional to the frequency up to 100 Mc/sec. This result is indicative of a relaxation. To verify the existence of such a relaxation, measurements of internal friction of single crystals have been made from 1.5°K to 300°K and in the 5- to 80-Mc/sec frequency region. The results for a 5-Mc/sec crystal are shown by Fig. 2.

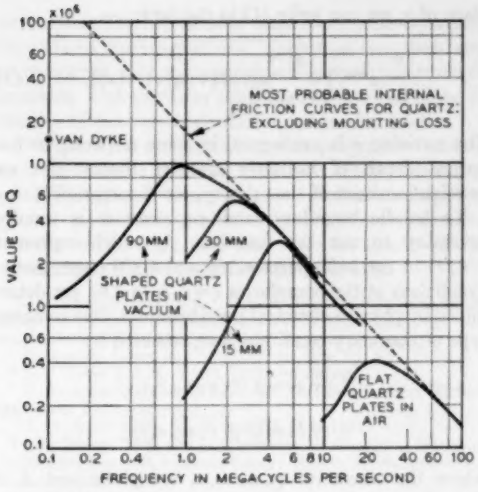


FIG. 1. Experimental measurements of Q of AT quartz crystals as a function of frequency and size. Dashed line indicates most probable internal friction as a function of frequency for quartz alone.

As can be seen, Q^{-1} increases as the temperature decreases, shows two maxima and then drops to a low value which is thought to be determined by the mounting loss. The internal friction, as shown by the dashed lines, can be divided into three parts labeled mounting loss, dislocation relaxation, and oxygen vibrations.

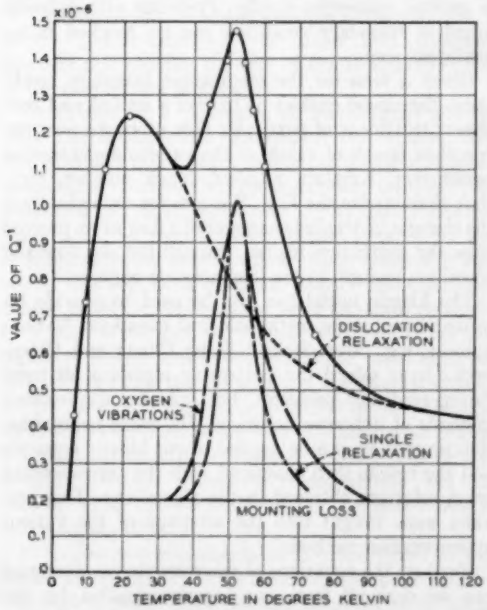


FIG. 2. Internal friction for a 5-Mc/sec AT quartz crystal as a function of temperature.

To justify this division, measurements were made for a number of crystals whose frequencies ranged from 5 to 80 Mc/sec. In every case two peaks were obtained and Fig. 3 shows a plot of the relaxation frequencies against $1/T$. There are two relaxation frequencies determined by the equations of Fig. 3. The high-energy curve agrees well with one measured previously in fused silica.[1] This peak is thought to be due to impurities distorting the lattice until it approximates the fused silica lattice in the neighborhood of the impurities. The height of the peak is from 1/2000 to 1/20 of the height of the fused silica peak and the relaxation spectrum is much narrower indicating only a small range of bond angles.

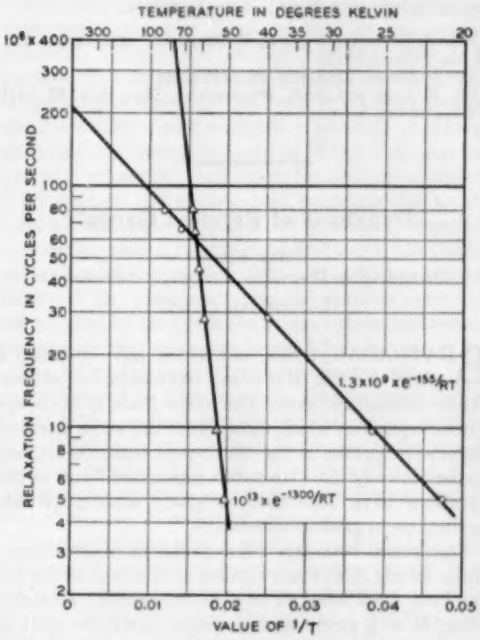


FIG. 3. Semilogarithmic plot of relaxation frequencies as a function of $1/T$.

The lower activation energy curve is thought to be caused by dislocations. The relaxation peak is similar to that found in metals.[2] This has been attributed to pinned dislocations being displaced from their minimum energy position by one atomic spacing in the glide plane. Another characteristic of dislocation loss is the exponential increase at high temperatures due to breakaway of dislocations from their impurity pinning points. This type of loss in quartz, occurring at high temperatures, is shown by the work of Cook and Breckenridge.[3] From these data and the theoretical equations of reference 2, one finds values for dislocation

number, loop length, and ratio of limiting shear stress to elastic shear modulus compared with a metal in Eq. (1).

	N dislocations per cm³	Loop length in cm	Ratio of $(T_{13})_0/\mu$	
Metal	10^8 to 10^9	4×10^{-4}	5×10^{-4}	(1)
Quartz	10^2	2×10^{-3}	4.5×10^{-5}	

The lower value of $(T_{13})_0/\mu$ indicates that the dislocations are broader than those in metals. The lower number of dislocations present accounts for the smaller dissipation and greater mechanical stability of quartz. The small residual aging of the elastic properties of quartz noted after processing the crystal is probably due to the stabilizing of dislocation positions as a function of time. This aging should be eliminated if the crystal is held at a very low temperature.

[1] See O. L. Anderson and H. E. Bömmel, J. Am. Ceram. Soc. **38**, No. 4 (April, 1955).
[2] W. P. Mason, Phys. Rev. **98**, 1136 (1955).
[3] R. K. Cook and R. G. Breckenridge, Phys. Rev. **92**, 1419 (1951); see Fig. 2.

Dynamics of Rarefied Gases*

MAX KROOK
Harvard College Observatory, Cambridge, Massachusetts

(Received July 21, 1955)

BHATNAGAR, Gross, and Krook[1] have proposed a model for the statistical representation of molecular collisions in gases. The model leads to nonlinear kinetic equations which incorporate the same essential physical properties as the Boltzmann equation but are mathematically more tractable. Linearized forms of the equations have been used to study small-amplitude oscillations in gaseous systems.[1,2]

For several boundary-value problems of one-dimensional steady flow, exact numerical solutions of the full nonlinear equations can be obtained without excessive labor. In such problems with simple gases, the state of a system is specified by a distribution function $f(\mathbf{v},x)$, where x = position coordinate and $\mathbf{v}\equiv(v_1,v_2,v_3)$ = molecular velocity. Let m = molecular mass, $n(x)$ = number density, $\mathbf{q}(x)$ = flow velocity, $T(x)$ = kinetic temperature, and $c(x)=[kT(x)/m]^{\frac{1}{2}}$ (k = Boltzmann's constant). In the absence of external forces, the kinetic equation is then[1]:

$$v_1\partial f/\partial x=-n\kappa f+n^2\kappa\Phi(\mathbf{v};\,\mathbf{q}(x),\,c(x)) \qquad (1)$$

with

$$\Phi(\mathbf{v};\,\mathbf{q},\,c)\equiv(2\pi c^2)^{-\frac{3}{2}}\exp[-(\mathbf{v}-\mathbf{q})^2/2c^2]; \qquad (2)$$

$n\kappa$ is the collision frequency; we have taken κ independent of $\mathbf{v}$ or equal to some average over $\mathbf{v}$. (Kinetic equations corresponding to more general models[2] can also be used in this work.)

Introducing $\tau=\int n\kappa dx$ as independent variable in place of x, we can write (1) in the form:

$$\frac{\partial}{\partial\tau}(e^{\tau/v_1}f)=\frac{e^{\tau/v_1}}{v_1}n(\tau)\Phi(\mathbf{v};\,\mathbf{q}(\tau),\,c(\tau)). \qquad (3)$$

The variable τ is analogous, in some respects, to the optical depth of radiative transfer theory (for an excellent account of that theory, see Kourganoff[3]).

To handle boundary-value problems it is usually necessary to use the functions $f_\pm$ which represent $f(\mathbf{v};\tau)$ in the half-spaces $v_1>0$ and $v_1<0$ respectively. Conditions at the boundaries $\tau=a,\,b$, $(a<b)$, are determined by physico-chemical considerations. One common type of boundary condition is represented by

$$\begin{aligned} f_+(\mathbf{v},a)&=A\cdot\Phi(\mathbf{v};\,\mathbf{q}_1,\,c_1),\\ f_-(\mathbf{v},b)&=B\cdot\Phi(\mathbf{v};\,\mathbf{q}_2,\,c_2), \end{aligned} \qquad (4)$$

where the constants $\mathbf{q}_1$, $\mathbf{q}_2$, c_1, c_2 are given and A, B are to be determined. Various other forms of boundary condition also have to be considered, e.g., specular reflection, etc.

Directly from (3), (4), we find explicit expressions for $f_\pm(\mathbf{v},\tau)$ as integrals involving n, $\mathbf{q}$, c and A, B. Multiplying these solutions successively by 1, $\mathbf{v}$, $(\mathbf{v}-\mathbf{q})^2$ and integrating with respect to velocity, we obtain a set of integral equations which suffice to determine the unknowns $n(\tau)$, $\mathbf{q}(\tau)$, $c(\tau)$, A, B, and hence also $f(\mathbf{v},\tau)$. The integral equations can be solved numerically by a process of successive approximations which, in general, converges rapidly. Problems with different forms of boundary condition can be handled in an analogous way.

Given a *form* for the microscopic boundary conditions, the model enables us to give a unified and consistent treatment of particular flow problems over the complete ranges of values of the various dimensionless parameters (Knudsen number, Mach number, etc.), that characterize the flow. We can, for example, trace the changes in detailed character of a flow as we proceed from the continuum regime, through the slip flow and transition regions, to the free molecule regime.

The kinetic model can also be used to provide information on some methodological questions. Various authors, e.g., Mott-Smith,[4] Wang Chang and Uhlenbeck,[5] have solved the Boltzmann equation for some "steep-gradient" problems, but only by approximate methods of unknown accuracy. The same approximation procedures can be applied to our kinetic equation and the results then compared with the corresponding exact solutions obtained in the above way. This provides some insight into the accuracy of the various approximation methods.

Further, the equations of macroscopic gas dynamics can be derived[2] from our kinetic equation by the Chapman-Enskog procedure; explicit values are thereby obtained for the coefficients of viscosity and of heat

conduction. The solutions of these continuum equations for various "steep-gradient" problems can then also be compared with the corresponding exact microscopic solutions. This provides some insight into the range of validity of the Navier-Stokes equations for gases.

Extensive calculations are now in progress on several such problems, e.g., shear flow, heat transfer between parallel plates, structure of shock fronts, atmosphere-meteor interaction problems, etc. A more detailed account of the work, together with the results of the calculations, will be published in due course.

* The research reported in this paper has been supported by the Geophysics Research Directorate, Air Force Cambridge Research Center, under a contract with Harvard University.

[1] Bhatnagar, Gross, and Krook, Phys. Rev. **94**, 511 (1954).
[2] E. P. Gross and M. Krook (to be published).
[3] V. Kourganoff, *Basic Methods in Transfer Problems* (Oxford University Press, Oxford, 1952).
[4] H. M. Mott-Smith, Phys. Rev. **82**, 885 (1951).
[5] C. S. Wang Chang and G. Uhlenbeck, University of Michigan Engineering Research Institute Report, Project M999, September, 1953 (unpublished).

Photoluminescent Modulation in Nonuniformly Excited ZnS Phosphors

R. E. Halsted
General Electric Research Laboratory, Schenectady, New York
(Received July 18, 1955)

PHOTOLUMINESCENT modulation, which is described below, provides a simple demonstration of *n*-type photoconductivity in some ZnS phosphor powders and clarifies the role of conduction electrons in the luminescent process.

Irradiation of a phosphor at wavelengths shorter than the fundamental absorption edge creates free electrons and holes which recombine preferentially at luminescent centers. Strong absorption confines this ionization to a small region near the irradiated surface. An alternating electric field applied parallel to the incident radiation moves mobile charge carriers in and out of this region. If electrons provide the dominant charge transport, luminescent recombination proceeds at a rate dependent on the electron density in the region of initial ionization. Maximum recombination rate occurs when the electrode on the irradiated side of the sample is becoming positive for a phosphor specimen so mounted between electrodes that no charge is introduced through its surface. If holes provide the dominant charge transport, recombination will be favored when the electrode on the irradiated side is becoming negative.

If luminescent recombination is proportional to the instantaneous density of the mobile carrier in the ionization region, pronounced photoluminescent modulation should occur. One would anticipate a limited variability in the phase and complexity in wave form of such modulation from its dependence on the details of charge motion. Nevertheless, the sign of the mobile carrier is unambiguously identified by the phase of photoluminescent modulation observed.

Such modulation has been observed with 2537 A excitation of ZnS:Ag, Cl and ZnS:Cu, Al; and 3650 A excitation of ZnCdS(35, 50, and 85% Cd):Ag, Cl and ZnSSe(20% Se):Cu, Cl phosphor powder samples. These were prepared by dispersing the phosphor in an organic film between plane electrodes—one transparent. The ultraviolet lamps were dc operated. The photoluminescence was observed with a photomultiplier and displayed on an oscilloscope. Filters were used to confine excitation and observation to the desired wavelengths.

Photoluminescent modulation with a fundamental frequency component equal to that of the applied sinusoidal voltage was observed in each case. The luminescent maxima occurred when the irradiated side of the samples was becoming positive or at its positive maximum. Detectable modulation occurred at average sinusoidal field strengths as low as 10 volts/cm. Irradiation at a wavelength longer than the absorption edge or application of the electric field perpendicular to the incident radiation produced the previously reported modulation with a fundamental frequency component twice that of the applied field and a higher voltage threshold for detection.[1] The observations were not complicated by changes in the average photoluminescent emission at field strengths below 10^3 volts/cm. At 10^3 volts/cm modulation amplitudes greater than 50%[2] were observed in some samples. Further examination of the 3650 A excited samples revealed a slow decrease of modulation amplitude with increasing excitation intensities between 0.1 and 10 microwatts/cm^2. The modulation amplitude increased slowly with increasing frequency from 20 to 20 000 cycles/sec.

The observations require electrons associated with the luminescent process to travel distances greater than activator separation. Luminescent emission must be strongly dependent upon the instantaneous free electron density in the excitation region. The weak frequency dependence suggests a ratio of electron to hole drift mobilities greater than 1000.

These observations appear incompatible with a recently proposed model for sulfide phosphors requiring emission by hole capture,[3] and limit possible separation of luminescent and volume photoconductive processes in these materials.[4,5]

The author gratefully acknowledges the helpful suggestions and criticisms of S. Roberts and F. E. Williams, and the assistance of H. S. Dunkerley with a portion of the experimental work.

[1] F. Matossi, Phys. Rev. **94**, 1151 (1954).
[2] Percent photoluminescent modulation = 100(maximum − minimum)/(maximum + minimum).
[3] J. Lambe and C. C. Klick, Phys. Rev. **98**, 909 (1955).
[4] R. H. Bube, Phys. Rev. **83**, 393 (1951).
[5] H. Kallman and B. Kramer, Phys. Rev. **87**, 91 (1952).

Frequency Dependence of Magnetocrystalline Anisotropy

R. M. Bozorth, Beatrice B. Ceilin, J. K. Galt, F. R. Merritt, and W. A. Yager
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received June 3, 1955)

IT is well known that the magnetocrystalline anisotropy constant K_1 can be determined both by static measurement of the torque on a single crystal placed in a strong magnetic field, and by measurement of the magnetic field for resonance at microwave frequencies. In experiments on several specimens of nickel-iron ferrites it was apparent that the same value of K_1 was not obtained by the two methods. In order to determine whether the difference was due to chemical composition or to a true frequency dependence, measurements were made by both methods on specimens cut from the same single crystal. The crystal was obtained from Dr. G. W. Clark of Linde Air Products Company; chemical analysis for Ni and Fe indicate that it has the composition $Ni_{0.76}Fe_{2.16}O_4$. The simplest formula preserving these proportions is $(NiO)_{0.70}Fe_2O_3$, which calls for metal ion vacancies in the lattice but no Fe^{++} ions. The electrical resistivity of this material at room temperature, however, is $\rho = 1$ ohm-cm, and we believe that this low value of resistivity indicates at least the temporary existence of Fe^{++} ions.

Our results, given in Table I, show that the difference in K_1 is slight at room temperature but marked at -196°C. This difference is the first time to our knowledge that frequency dependence has been observed in a magnetocrystalline anisotropy constant.

TABLE I. Magnetic anisotropy constant K_1 in ergs/cm³, and anisotropy field $2K_1/M_s$ in oersteds, as dependent on frequency and temperature.

Quantity	Temperature	$Ni_{0.8}Fe_{2.2}O_4$ Static	$Ni_{0.8}Fe_{2.2}O_4$ Microwave	$Mn_{1.8}Fe_{1.2}O_4$ Static	$Mn_{1.8}Fe_{1.2}O_4$ Microwave
K_1	20°C	-39×10^3	-43×10^3	-34×10^3	-33×10^3
K_1	-196°C	-42×10^3	-74×10^3	-240×10^3	-233×10^3
$2K_1/M_s$	20°C	-260	-280	-160	-175
$2K_1/M_s$	-196°C	-260	-468	-840	-812

The conclusion that Fe^{++} ions are present leads us to believe that this variation of K_1 with frequency confirms the interpretation which some of us have given[1] to relaxation phenomena observed in other crystals where Fe^{++} ions were certainly present. This interpretation starts with the assumption that there are Fe^{++} and Fe^{+++} ions in the crystal on which the valence electrons rearrange easily with a short time constant as suggested some time ago by Verwey.[2] It is then suggested that these electronic rearrangements occur as the magnetization changes direction, and affect the magnetocrystalline anisotropy constant, so that the value of K_1 will differ depending on whether it is measured at an angular frequency above or below the reciprocal of the time constant for the electronic rearrangement. The microwave measurements indicate that the reciprocal of this time constant corresponds to 24 000 Mc/sec at about 160°K.

A similar comparison of static and microwave measurements of magnetocrystalline anisotropy on a single crystal of manganese ferrite of composition $Mn_{0.98}Fe_{1.86}O_4$, was made using our static value and the dynamic values deduced by S. Geschwind and J. F. Dillon, Jr. from their microwave measurements at 20° and -196°C, at frequencies of 9300 and 48 000 Mc/sec. These values, also given in Table I, show that K_1 is practically independent of frequency in this material. This indicates that electrons are not free to move so easily from one atom to another in this material as compared with the nickel-iron ferrite.

We are indebted to S. Geschwind and J. F. Dillon, Jr., for permission to quote their results before publication, and to F. J. Morin for the measurement of resistivity.

[1] Galt, Yager, and Merritt, Phys. Rev. **93**, 1119 (1954); J. K. Galt, Bell System Tech. J. **33**, 1023 (1954).
[2] E. J. W. Verwey, Nature **144**, 327 (1939).

Emission Spectroscopy at Liquid Helium Temperature*

Joseph G. Hirschberg
University of Wisconsin, Madison, Wisconsin
(Received August 1, 1955)

A KOVAR and glass hollow-cathode discharge tube has been operated immersed in liquid helium, to study the emission spectrum of helium. The spectrum was recorded directly from an RCA 1*P*21 photomultiplier, with a Perot-Fabry interferometer and a 21-foot stigmatically mounted concave grating swept in synchronism.[1,2] The grating was swept by moving the exit slit, and the Perot-Fabry by varying the interplate gas pressure linearly with time. The entrance and exit slits were curved to concentrate the light into a single interference fringe. Details will be published after further work in improving the signal-to-noise ratio.

The line $\sigma = 21\,335$ K ($\lambda = 4686$ A) of He II was observed, and the Doppler broadening was found to be sensibly reduced from the amount observed with liquid-air cooling. The splitting arising from the radiation shift[3] was clearly shown. In particular, the $3s_{\frac{1}{2}}-4p_{\frac{1}{2}}$ and $3p_{\frac{1}{2}}-4s_{\frac{1}{2}}$ pair was almost completely resolved and the $3s_{\frac{1}{2}}-4p_{\frac{3}{2}}$ and $3p_{\frac{3}{2}}-4d_{\frac{3}{2}}$ pair was partly resolved.

It was found that a discharge could be not only maintained at this low temperature, but easily started as well. In all, about three liters of liquid helium were consumed in 25 minutes' operation of the discharge. The current averaged 30 milliamperes and the potential difference 250 volts.

It is concluded that liquid helium can certainly be used to advantage to cool a helium discharge in cases where Doppler broadening presents a serious problem at liquid air or even liquid hydrogen temperatures, and that no serious experimental difficulties are experienced.

It is expected that by using helium as a carrier gas, the spectra of other elements can be obtained at liquid helium temperature as readily as at liquid air temperature.

This work was suggested by Professor J. E. Mack. The liquid helium was very kindly supplied by Professor J. R. Dillinger and his group. An especially sensitive photomultiplier was lent by Professor A. D. Code of the department of astronomy. The Perot-Fabry plates were multilayer-coated for σ 21 335 at Bellevue laboratory through the courtesy of Professor P. Jacquinot, with whom I have had the privilege of discussing high-resolution techniques.

* Supported by the Research Committee of the University and by the Office of Ordnance Research.

[1] J. G. Hirschberg, Phys. Rev. **99**, 623(A) (1955).

[2] Since our initial use of the sweeping system our attention has been called to the suggestion of a somewhat analogous system by J. H. Jaffe, Nature **168**, 381 (1951).

[3] J. E. Mack and N. Austern, Phys. Rev. **72**, 972 (1947).

Nature of the Water-Vapor-Induced Excess Current on Grown Germanium *p-n* Junctions

EDWARD N. CLARKE

Physics Laboratory, Sylvania Electric Products, Bayside, New York

(Received July 11, 1955)

IT is well known that water vapor can induce large excess currents on germanium *p-n* junctions biased in the reverse direction. What is not well understood or agreed upon at present is the nature of the water-vapor-induced currents. In general, the excess current might be either ionic or electronic or possibly a combination of the two. Law[1,2] has presented evidence in favor of the ionic interpretation, while Christensen[3] and McWhorter and Kingston[4] have presented data which seem to favor an electronic interpretation. Electronic currents are associated with the presence of an electron-conducting layer or "channel" on the *p*-type germanium.

In the present work, a new criterion has been used to determine the nature of the water-vapor-induced current. This criterion involves comparison of two curves, namely, (1) the water-vapor-induced excess current-voltage curve and (2) a field-induced excess current-voltage curve. Such a comparison is shown in Fig. 1 and described below. For the field-induced channel, a positive electric field applied perpendicular to the surface attracts electrons to the surface inducing an electron conducting layer as shown in Fig. 2. The field-induced experiment is carried out in air in the absence of water vapor, and control of the channel conductance is achieved by means of the external inducing field. The present field experiment is carried out at 207°K because such a dc field-induced conducting layer is stable only at low temperatures.[5] It is sometimes necessary to use even lower temperatures to stabilize the channel. In such an experiment, the possibility of ionic conduction through a thin surface film of water does not exist and it can then be concluded that the excess current flowing across the junction is wholly electronic in nature. If the water-vapor-induced excess current-voltage curve and the field-induced excess current-voltage curve are identical in shape as shown in Fig. 1, we shall conclude that the two mechanisms are likewise identical and hence that the water-vapor-induced current is electronic rather than ionic.

The results shown in Fig. 1 were obtained on a grown germanium *p-n* junction having the following characteristics: approximate *p*-type resistivity 10 ohm-cm,

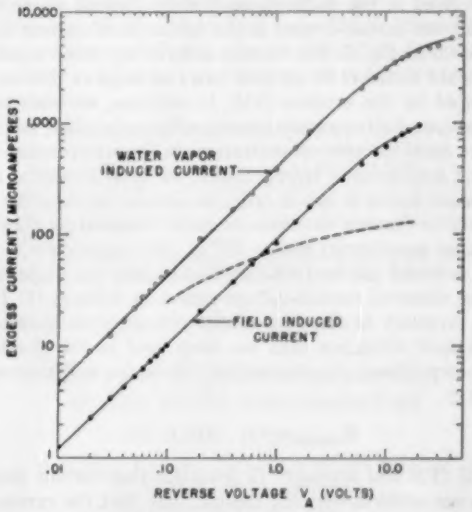


FIG. 1. Water-vapor-induced and field-induced excess current on the same germanium *p-n* junction. Water vapor pressure corresponds to about 100% relative humidity. Field-inducing voltage has been chosen so that field-induced current is four times smaller than water-induced current. Dashed line shows current-voltage dependence expected from simple channel theory.

n-type resistivity 1 ohm-cm, and junction circumference 1 cm. This is a particularly interesting case in which the water-vapor-induced excess current-voltage curve is not described by simple channel theory[4] which provides a current-voltage curve with shape shown by the dashed line. The measured excess current rises much more steeply with increasing voltage than predicted by simple channel theory. Nevertheless, the water-vapor-induced current in this case is electronic rather than

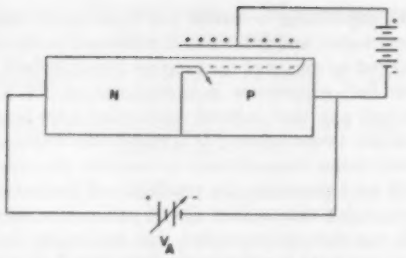


FIG. 2. Field-induced electron conducting channel on germanium p-n junction. A thin piece of mica separates the electrode and the germanium surface. Current is assumed to flow along channel and then through spots of low resistance as shown by arrow.

ionic which we conclude from the fact that the water-induced and field-induced curves have the same shape when compared, as shown, separated by a factor of four in magnitude of current. The factor of four results from the following considerations. There are two surfaces involved in the water-vapor-induced current, whereas only one surface is used in the field-induced current as shown in Fig. 2. The current induced by water vapor should therefore be at least twice as large as that induced by the external field. In addition, we wish to compare the two curves under conditions in which there are equal electron concentrations in the water-induced and field-induced layers. Hence, we must introduce a second factor of two in order to account for the difference in electron mobilities at room temperature (H_2O vapor experiment) and at 207°K (field experiment).

A model can be formulated to explain the shape of the observed current-voltage curves as follows: (1) It is necessary to assume a smaller voltage dependence of channel resistance than has been used in the simple theory (direct proportionality). Thus, we will assume that

$$R_{\text{channel}} \propto V^{m}. \quad (0 \leqslant m < 1)$$

(2) It is also necessary to postulate that current flow is not uniform over the channel, but that the current flows through the conducting layer and then through a spot (or line, or many such spots along the junction) into the bulk of the crystal as shown by the arrow in Fig. 2. Such spots might be regions of low resistance. With such a model it is easy to show that excess current

$$I \propto V_A^{1-m}$$

if it is assumed that most of the voltage drop occurs in the conducting surface layer rather than through the conducting spot. For the curves shown in Fig. 1, $m \approx 0$ over a wide range of voltages. At voltages above 4 or 5 volts, m appears to become an increasing function of applied voltage, a result to be expected if an appreciable number of conducting electrons are removed from the channel which at lower voltage is known to contain about 10^{12} conducting electrons per cm² of surface. The above expression for excess current is then not valid at the higher voltages.

[1] J. T. Law, Proc. Inst. Radio Engrs. **42**, 1367 (1954).
[2] J. T. Law and P. S. Meigs, Spring Meeting of the Electrochemical Society, Cincinnati, May, 1955 (unpublished).
[3] H. Christensen, Proc. Inst. Radio Engrs. **42**, 1371 (1954).
[4] A. L. McWhorter and R. H. Kingston, Proc. Inst. Radio Engrs. **42**, 1376 (1954).
[5] S. R. Morrison, Phys. Rev. **99**, 1655(A) (1955).

X-Ray Study of Phase Transition of Ferroelectric $PbTiO_3$ at Low Temperature

JINZO KOBAYASHI AND RYUZO UEDA
Department of Applied Physics, Waseda University, Tokyo, Japan
(Received June 27, 1955)

LEAD titanate, $PbTiO_3$, is a ferroelectric of particular interest in two respects. Firstly, $PbTiO_3$ has polar symmetry based on a monomolecular unit cell in the ferroelectric region. By contrast, other perovskite-type $PbMe^{+4}O_3$ crystals have nonpolar puckered structures with multiplications of perovskite cell sides in their room temperature modifications; $PbZrO_3$,[1] $PbHfO_3$,[2] and $PbSnO_3$ [3] are cases in point. Secondly, structural evidence of markedly large spontaneous polarization in the tetragonal modification has recently been discovered by Shirane, Pepinsky, and Frazer[4] and by the present authors[5]: metal ions undergo large shifts with respect to the oxygen framework. This cannot be understood on the basis of ionic radii, since Zachariasen's tolerance factor is nearly unity. The similar compound $SrTiO_3$, where the radius of Sr^{+2}(1.13 A) lies near that of Pb^{+2}, remains cubic down to absolute zero. It

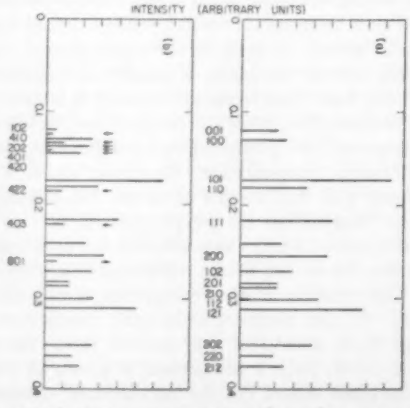


FIG. 1. Schematic representations of powder patterns. Abscissas indicate $\sin\theta/\lambda$; ordinates indicate uncorrected intensities. (a) Original perovskite structure. (b) Superstructure, where large arrows indicate the superstructure lines.

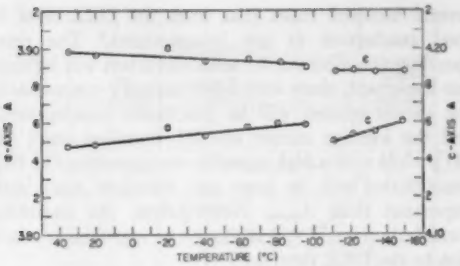


FIG. 2. Lattice spacing as a function of temperature.

appeared worthwhile to re-examine the structure of $PbTiO_3$ at lower temperature, although no phase transition has been found heretofore except for the ferroelectric one at the Curie point, 490°C.[4]

We have carried out an x-ray study of a powder specimen cooled by liquid air. $PbTiO_3$ crystals were grown from the melt, using stoichiometric mixtures of PbO and TiO_2 and $PbCl_2$ as flux; the powder specimen was prepared from this material.

Thermal hysteresis of the substance was found to be very large, and we found no evidence of a low-temperature phase change in a specimen cooled suddenly to −150°C. However, the specimen under gradual cooling showed a definite transition at about −100°C. X-ray powder photographs of an hour exposure were taken at steps of 30°C down from room temperature. We repeated the above procedures and found that the transitions were reproducible except for a slight scattering of of transition temperatures for different samples and cooling rates.

At about −100°C, extra lines appear. An important feature is that the extra lines appear only within the region of small values of $\sin\theta/\lambda$, i.e., 0.26×10^8 A^{-1}; and no change in intensity of diffraction could be observed in high-angle spectra before and after the transition. Furthermore, the new lines appear not as splitting due to lowering of crystal symmetry, but rather as superlattice lines, indices being assignable by assuming that $a'=4a$ and $c'=2c$, where a and c are the former perovskite axes. The observed spectra are indicated in Fig. 1, where ordinates show uncorrected intensities.

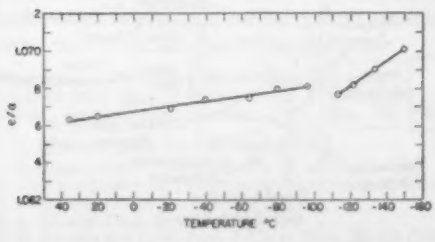


FIG. 3. Axial ratio as a function of temperature.

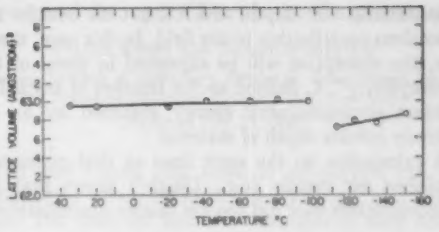


FIG. 4. Unit cell volume as a function of temperature.

The plotting of lattice constants and axial ratio *vs* temperature are shown in Figs. 2–4. The temperature dependence of the c-axial length is notable. Its expansion coefficient is negative, and the behavior is more remarkable after the transition in which a volume contraction occurs. It is to be stressed that the steep rise in axial ratio is observed with decreasing temperature between the transition point and −150°C.

It is possible that ferroelectric $PbTiO_3$ transforms into an antiferroelectric structure at about −100°C. However, other measurements and a detailed investigation of the crystal structure are needed to determine whether this is the case. Observations to these ends are in progress.

[1] Sawaguchi, Maniwa, and Hoshino, Phys. Rev. **83**, 1078 (1951).
[2] G. Shirane and R. Pepinsky, Phys. Rev. **91**, 812 (1953).
[3] Naray Szabo, Műegyetemi Kozlemények **1**, 30 (1947).
[4] Shirane, Pepinsky, and Frazer, Phys. Rev. **97**, 1179 (1955).
[5] J. Kobayashi and R. Ueda, lecture given at the annual meeting of the Physical Society of Japan, October, 1954 (to be published).
[6] G. Shirane and S. Hoshino, J. Phys. Soc. Japan **6**, 265 (1951); S. Nomura and S. Sawada, J. Phys. Soc. Japan **10**, 108 (1955).

Theory of the Infrared Absorption by Carriers in Semiconductors

R. B. DINGLE
Department of Physics, University of Western Australia, Nedlands, Western Australia
(Received July 19, 1955)

IN a recent paper Kahn[1] has shown that the observed magnitudes of infrared absorption by n-type germanium and p-type silicon are approximately those calculated from the Drude-Kronig-Zener (DKZ) formula when the effective carrier masses inserted in the theory are those obtained from diamagnetic ("cyclotron") resonance experiments.

It should be noted, however, that the DKZ absorption should be supplemented by a term arising from the anomalous nature of the skin effect.[2–5] It is not possible to express this correction in the form of a modification to the "absorption coefficient" quoted by Kahn[1] from Seitz,[6] since the definition of this quantity presupposes that the electric field decays precisely

exponentially with depth, which is not the case for the anomalous contribution to the field. In this note, therefore, the absorption will be expressed in terms of the "absorptivity" A, defined as the fraction of normally-incident electromagnetic energy absorbed by an effectively infinite depth of material.

A calculation on the same lines as that previously published for metals (e.g., Dingle[3]) shows that for semiconductors with not too low carrier concentrations:

$$A=\left(\frac{3\omega\hbar^3}{2\pi^2m^2e^2}\right)^{\frac{1}{3}}\mathrm{Re}\left[\left[\frac{i}{\mathfrak{U}\left(\frac{\tau v}{1+i\omega\tau}\right)+\frac{3i\mathfrak{p}\omega\hbar^3}{32\pi^3m^2e^2}}\right]^{\frac{1}{3}}\right]$$
$$+\tfrac{3}{4}(1-\mathfrak{p})\frac{\omega}{c}\mathrm{Re}\left[\frac{i\mathfrak{U}\left[\left(\frac{\tau v}{1+i\omega\tau}\right)^2\right]}{\mathfrak{U}\left(\frac{\tau v}{1+i\omega\tau}\right)}\right], \quad (1)$$

where Re stands for "the real part," $\mathfrak{U}$ for the integration operator $\mathfrak{U}\equiv(-\int_0^\infty dv\,v\frac{\partial f_0}{\partial v^2})$ with f_0 = Fermi-Dirac distribution function and v = carrier velocity, ω = circular frequency, m = effective carrier mass, τ = relaxation time, $\mathfrak{p}$ = fraction of carriers specularly reflected at the internal surface, and $\mathfrak{p}$ is the quantity defined in Dingle[2] taking into account the effects of displacement current, atomic polarization, and the presence of absorption bands near the frequency region investigated. The first term in (1) is that predicted by the DKZ theory, and the second that arising from the anomalous nature of the skin effect.

With the approximations (a) $\mathfrak{p}=0$, as found for metals,[2] (b) combined effect of displacement current, atomic polarization and absorption bands negligible, a supposition normally valid for frequencies in the infrared (except of course very close to absorption bands), and (c) $\omega\tau\gg1$, a condition usually satisfied for frequencies above the far infrared, the second (anomalous) contribution to (1) reduces to

$$A_{\mathrm{anom}}\simeq\left(\frac{2kT}{mc^2}\right)^{\frac{1}{3}}\frac{F_1(T_0/T)}{F_{\frac{1}{2}}(T_0/T)}, \quad (2)$$

where the F's are Fermi-Dirac integrals[7,8] and T_0 is the degeneracy temperature.

In the degenerate limit $T\ll T_0$, as for metals

$$A_{\mathrm{anom}}\to\tfrac{3}{4}v_{\mathrm{Fermi}}/c; \quad (3)$$

while in the nondegenerate limit $T\gg T_0$,

$$A_{\mathrm{anom}}\to\left(\frac{8kT}{\pi mc^2}\right)^{\frac{1}{3}}\sim\frac{\text{thermal velocity of carriers}}{\text{velocity of light in vacuo}}. \quad (4)$$

For metals, this anomalous contribution to A may be several hundred times that from the DKZ term for good conductors at low temperatures.[2] The corresponding contribution for semiconductors will be much less important, since with a *low* impurity concentration the carrier system will be practically nondegenerate and the average carrier velocity therefore small [see (4)], while with a *high* impurity concentration the DKZ contribution will be large and therefore much more important than A_{anom}. Nevertheless, the anomalous contribution could still represent an appreciable correction to the DKZ theory.

[1] A. H. Kahn, Phys. Rev. **97**, 1647 (1955).
[2] R. B. Dingle, Physica **18**, 985 (1952); **19**, 311, 348 (1953).
[3] R. B. Dingle, Physica **19**, 729 (1953).
[4] T. Holstein, Phys. Rev. **88**, 1427 (1952).
[5] R. Wolfe, Proc. Phys. Soc. (London) **A68**, 121 (1955).
[6] F. Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), pp. 630–640.
[7] J. McDougall and E. C. Stoner, Phil. Trans. **A237**, 67 (1938).
[8] P. Rhodes, Proc. Roy. Soc. (London) **A204**, 396 (1950).

Optical Absorption Spectrum of MgO by Reflectivity*

JAMES R. NELSON

Department of Physics, Cornell University, Ithaca, New York

(Received August 1, 1955)

IN the course of work to be published on absorption spectra of NaCl and KCl at 300°K and at 140°K by a reflectivity method,[1–3] a measurement was also made on a large single crystal of MgO obtained from the Norton Company (Canada). The apparatus for making this measurement appearing in Fig. 1 will be presented in greater detail in the forthcoming paper on alkali halides.

The result of this measurement on MgO is shown in Fig. 2. In discussing it recently,[4] Krumhansl has pointed out that he believes the narrowing of the 1640 A peak at 140°K is fairly good evidence of exciton absorption. Also, we feel from our absorption studies of NaCl and KCl both by transmission and reflectivity, that this

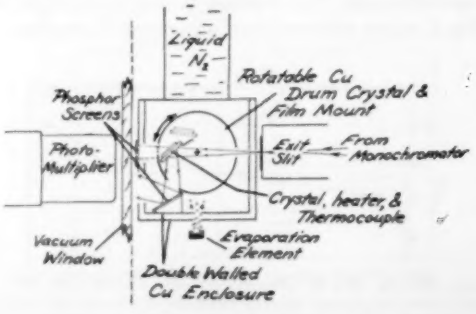


FIG. 1. Diagram of apparatus.

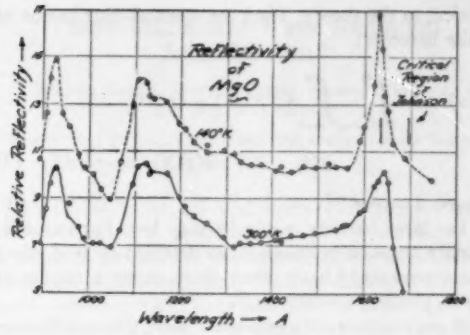


FIG. 2. Relative reflectivity of MgO as a function of wavelength.

peak is a truly fundamental absorption band with absorption coefficients of the order of 10^6 cm^{-1}, and not weak absorption due to impurity levels. This peak, indeed, confirms the results of Johnson[5] who made fluorescence and transmission measurements on thin crystals of MgO of known impurity content and also on samples of lower impurity content from the University of Missouri. Johnson found that in the region (indicated in Fig. 2) from 1635 A to 1695 A there is a rapid change in the excitation spectrum for fluorescence, and that at 1695 A the absorption coefficient is rising very rapidly. From these results, he concluded that there must be fundamental absorption in this region.

Before considering the shorter wavelength peaks, it should be pointed out that there is no resolvable splitting in the exciton peak of MgO at low temperature as we have observed in NaCl and KCl and as has also been seen in BaO.[6]

The temperature-independent peaks at 1120 A and 930 A are tentatively interpreted as absorption due to transitions from the uppermost filled band to the normally empty conduction band. This absorption band begins at about 1240 A; the band gap, then, would be about 10 ev and not 5.9 ev as reported recently by Saksena and Pant[7] who made studies on natural crystals of MgO.

In closing, it is interesting to compare this 10-ev band gap value with the approximate 11-ev value that Lempicki[8] gives as the threshold primary electron energy for secondary emission in MgO. This may be confirmation of his original supposition that the secondary electrons come from the uppermost filled band.

* Supported in part by the Office of Naval Research.
[1] Frederick Seitz, Revs. Modern Phys. **26**, 27 (1954).
[2] Nelson, Siegfried, and Hartman, Phys. Rev. **99**, 658(A) (1955).
[3] P. L. Hartman, Phys. Rev. **99**, 658(A) (1955).
[4] J. A. Krumhansl, "Photoexcitation in ionic crystals," Atlantic City Conference on Photoconductivity (to be published).
[5] P. D. Johnson, Phys. Rev. **94**, 845 (1954).
[6] R. J. Zollweg, Phys. Rev. **97**, 288 (1955).
[7] B. D. Saksena and L. M. Pant, J. Chem. Phys. **23**, 989 (1955).
[8] A. Lempicki, Proc. Phys. Soc. (London) **B66**, 278 (1953).

Theory of Acceptor Levels in Germanium*

WALTER KOHN† AND DANIEL SCHECHTER†
Carnegie Institute of Technology, Pittsburgh, Pennsylvania and Bell Telephone Laboratories, Murry Hill, New Jersey
(Received July 25, 1955)

IT is well known that holes in germanium can be bound by Group III impurities in so-called acceptor levels. The observed ionization energies of such holes are 0.0104 ev (B),[1,2] 0.0102 ev (Al),[2] 0.0108 ev (Ga),[2] and 0.0112 ev (In).[2] The fact that these energies differ by only 10% suggests that the binding is largely due to the long-range Coulomb potential of the acceptors rather than to the less well understood forces in their immediate vicinity.

We have therefore carried out a theoretical calculation using a Coulomb potential[3] $-e^2/\kappa r$ and "mass"-parameters determined by cyclotron resonance experiments.[4] The resulting theoretical ionization energy is 0.0089 ev, in rather good agreement with experiment.

The calculation was based on the coupled effective mass equations which have recently been shown to arise when the band structure is degenerate.[6-7] In view of the fact that the spin-orbit splitting at the top of the valence band in Ge is about 30 times as large as the ionization energy of the acceptors,[8] the four coupled equations (V.15) of reference 7 are a good first approximation. Group-theoretical considerations suggest a trial function of the form (see reference 7):

$$ae^{-r/r_1}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}+be^{-r/r_2}\begin{bmatrix}z^2-\frac{1}{2}(x^2+y^2)\\0\\-(\frac{3}{2})^{\frac{1}{2}}(x^2-y^2)\\0\end{bmatrix}+ice^{-r/r_2}\begin{bmatrix}0\\(x+iy)z\\xy\\0\end{bmatrix}.$$

The parameters were varied to maximize the ionization energy. The value 0.00883 ev was obtained with the following parameters: $r_1=43.3\times10^{-8}$ cm, $r_2=33.8\times10^{-8}$ cm, $a=1.71\times10^{9}$ cm$^{-\frac{1}{2}}$, $b=-2.29\times10^{21}$ cm$^{-7/2}$, $c=4.97\times10^{21}$ cm$^{-7/2}$.

With this wave function as a starting point, the original six coupled equations (see reference 7, V.13) were treated by a perturbation-variation method, which increased the ionization energy by about 1% to 0.00893 ev.

A variational calculation, such as the present one, always leads to a low ionization energy. We estimate that the true eigenvalue of the six coupled equations lies in the range 0.0094 ± 0.0005 ev. The remaining discrepancies with the experimental values may be ascribable partly to the breakdown of the theory in the immediate vicinity of the acceptor ions and partly to inaccuracies of the mass parameters used.

Kittel and Mitchell[6] have reported a theoretical value of 0.022 ev for the ionization energy. The large discrepancy with our result and with experiment is due to an incorrect transformation of the Hamiltonian operator (see their concluding sentence). A better order of magnitude estimate is obtained from a simple

Bohr model with an effective mass $m^*=0.34m$ derived from the mean curvature of the "heavy" hole band. This gives an ionization energy of 0.018 ev.

Calculations of the acceptor levels in silicon are in progress.

* Work supported in part by the Office of Naval Research.
† Permanent address: Carnegie Institute of Technology, Pittsburgh, Pennsylvania.
[1] P. Debye, Phys. Rev. **91**, 208 (1953).
[2] T. Geballe and F. J. Morin, Phys. Rev. **95**, 1085 (1954).
[3] κ is the static dielectric constant which was taken to be 16.0; see H. B. Briggs, Phys. Rev. **77**, 287 (1950).
[4] Dresselhaus, Kip, and Kittel, Phys. Rev. **77**, 287 (1950). B. Lax *et al.*, Phys. Rev. **93**, 1418 (1954); for the present calculation we have used the values $A=-13.0\hbar^2/2m$, $|B|=8.7\hbar^2/2m$, and $|C|=11.4\hbar^2/2m$ which were kindly communicated to us by Dr. Dexter and Dr. Zeiger of M.I.T.
[5] W. Kohn and J. M. Luttinger, Phys. Rev. **96**, 529 (1954).
[6] C. Kittel and A. M. Mitchell, Phys. Rev. **96**, 1488 (1954).
[7] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[8] H. A. Kahn, Phys. Rev. **97**, 1647 (1955).

Generation of $1/f$ Noise by Levels in a Linear or Planar Array

S. Roy Morrison

Physics Laboratory, Sylvania Electric Products, Bayside, New York

(Received June 9, 1955)

TRAPPING levels in a linear or planar array (at edge dislocations or the surface), or agglomerates of levels, will possess the following important property. A trapped charge will cause a potential barrier to further trapping of like charges, and a fluctuation in the trapped charge will produce a proportional fluctuation in the barrier height (although the energy of ionization from the levels will be constant). We will show that trapping with this simple barrier property may lead to $1/f$ noise.

This barrier property has been associated with the Elovich equation[1-3] which describes in numerous cases the rate of irreversible adsorption. Thus, the application of the model to the many apparently disconnected cases of $1/f$ noise is possible.

Denoting by N the trapped charge in excess of equilibrium, we obtain from an analysis of the model:

$$dN/dt=B(e^{\beta N}-1), \tag{1}$$

where β is related to the "capacitance" between the levels and the bulk material, B represents the equilibrium rate at which charge crosses the barrier, and is very sensitive to temperature. The first term is thus the trapping rate, the second the ionization rate.

From Eq. (1),

$$N=-\beta^{-1}\ln(1-Ae^{-\gamma t}), \tag{2}$$

with A the constant of integration, $\gamma=-\beta B$.

If one assumes that the distribution $g(N)$ of N over the levels is Gaussian, with mean $\bar{N}$ and standard deviation ξ, the autocorrelation function can be calculated. The use of this distribution function is not critical to the theory. Then the spectral distribution of noise becomes:

$$G(\omega)=-4\int_{N=-\infty}^{\infty}\int_{\tau=0}^{\infty}N\beta^{-1}\ln\{1-Ae^{-\gamma\tau}\}\times g(N)\cos\omega\tau d\tau dN, \tag{3}$$

where $A=1-e^{-\beta N}$.

For large barriers, γ can be very low. For example, surface traps on germanium, as detected by field effect measurements,[4,5] have decay times about a minute at 20°C, yielding $\gamma=10^{-2}$ sec^{-1}.

If one assumes $\omega\gg\gamma$ as a lower limit, the contribution to $G(\omega)$ from the negative values of N turns out to be small, as the decay rate is slow.

For positive N, when the decay is fast, the logarithmic term in (3) can be expanded and the integration over t performed. Replacing the resulting summation by an integration, thus neglecting terms in γ/ω^2, and using the rough approximation that $\mathrm{Ci}(x)\sin x-\mathrm{Si}(x)\cos x+\frac{1}{2}\pi\cos x$ is a step function, zero for $x>1$ and $\pi/2$ for $x<1$, we obtain

$$G(\omega)=\tfrac{1}{2}\pi(\beta\omega)^{-1}\left\{\int_0^{\infty}Ng(N)dN-\int_0^{\beta^{-1}\ln(\omega/\gamma)}N(2\pi\xi^2)^{-\frac{1}{2}}\times\exp[-(N-\bar{N})^2/2\xi^2]dN\right\}; \tag{4}$$

and if $N-\bar{N}\ll\sqrt{2}\xi$ the second integral is a slowly varying function of ω and the $1/\omega$ distribution of noise is obtained. The requirement sets an upper limit on the $1/\omega$ spectrum, namely $\ln(\omega/\gamma)<\beta(\sqrt{2}\xi+\bar{N})$. It is interesting that the temperature-sensitive quantity γ does not appear in the expression for $G(\omega)$, except in the insensitive logarithmic form.

We have here developed the frequency spectrum of the trapped charge. Noise will appear in bulk conductivity measurements since charge trapped represents a decrease in current carriers. In contact or rectifier studies, the noise may arise from fluctuations in the barrier height; for small fluctuations the barrier height is proportional to the trapped charge. Thus the concepts presented may be extended to carbon contact devices and metal films, as well as semiconducting devices.

A more detailed discussion of the model and the experimental results on the field effect which led to this analysis will be published later.

The author is indebted to Professor H. Callen and Professor P. H. Miller, Jr., of the University of Pennsylvania for their helpful suggestions.

[1] H. A. Taylor and N. Thon, J. Am. Chem. Soc. **74**, 4169 (1952).
[2] D. Melnick, Thesis, Physics Department, University of Pennsylvania, 1954 (unpublished).
[3] S. R. Morrison, Advances in Catalysis **7**, 62 (1955).
[4] R. H. Kingston and A. L. McWhorter, Phys. Rev. **98**, 1191(A) (1955).
[5] J. Bardeen and S. R. Morrison, Physica **20**, 873 (1954).

Three-Millimeter Wave Radiation from the Sun*

WALTER GORDY, STEVE J. DITTO, JAMES H. WYMAN, AND ROY S. ANDERSON

Department of Physics, Duke University, Durham, North Carolina

(Received July 1, 1955)

BETWEEN the 2.5- and 5-mm wave oxygen resonances it is apparent that there should be a window in the earth's atmosphere which might allow detection of solar radiation. We have developed a 3-mm wave radiometer and have succeeded in detecting solar radiation at 3.2-mm wavelength. Under favorable weather conditions, a signal-to-noise ratio of about 25 to 1 could be obtained. Figure 1 shows a typical signal obtained by directing the antenna on and off the sun. Figure 2 shows, on a less favorable day, the results of allowing the sun to pass across the fixed antenna.

The antenna was an adapted (18-inch) searchlight reflector with a wave-guide feed. It is far from the optimum system that can be made with more time and money. The object of the present experiment was to find out if such additional expense and effort would be rewarding. The answer seems to be in the affirmative, especially if a higher and drier location than Durham, North Carolina, is chosen.

In the past the principal difficulty in attempting solar radiation studies in the 3-mm region has been the lack of adequate local oscillator power to produce a detectable beat note with the solar radiation. The present study shows that the *K*-to-*G* crystal harmonic generator developed by King and Gordy[1] is adequate but not optimum.

The Duke crystal multiplier assembly[1] was used both for the local oscillator and for the mixer of the superheterodyne receiver. To make this adaption, it was necessary only to feed the solar energy through the smaller guide onto the multiplier crystal and to connect the input of the i.f. amplifier directly across the tungsten whisker and crystal assembly.[2] The band width of the i.f. amplifier was 3 Mc/sec. The radiation chopper consisted of an attenuating strip which was pushed into and out of a wave-guide slot by the driving mechanism of a small radio speaker operated by an audio oscillator. The other features of the system were essentially those of the Dicke[3] microwave radiometer.

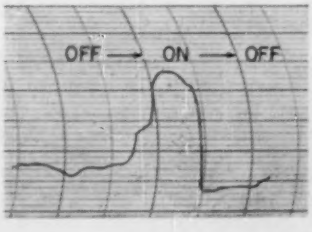


FIG. 1. Recorder signal of the sun at 3.20-mm wavelength obtained by directing the antenna toward and away from the sun.

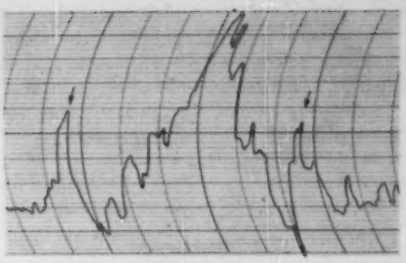


FIG. 2. Recorder tracing of the solar energy at a wavelength of 3.20 mm obtained by allowing the sun to pass over a fixed antenna. The arrows indicate the central image and two antenna side lobes of the sun's energy.

With a radiometer which is capable of detecting solar energy in the shorter millimeter wave region, the earth's atmosphere as well as the sun can be studied. Frequencies of 100 kMc/sec originate much deeper in the chromosphere[4] than do the 8-mm and longer radio waves now being used for solar investigations. It is possible that discrete spectral transitions of certain light molecules or radicals of the sun's photosphere might appear as absorption lines in the shorter millimeter region.

Propagation studies in the 3-mm region should have practical significance. The decrease in signal strength caused by haze or clouds was readily noticeable in these preliminary observations. It appears that a 3-mm wave radiometer might be of use in weather predictions. Ozone, a constituent of the upper atmosphere, is known to have intense rotational lines in the 3-mm wave region.

* This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

[1] W. C. King and W. Gordy, Phys. Rev. **90**, 319 (1953); **93**, 407 (1954).

[2] C. M. Johnson (private communication) has also used this adaptation for superheterodyne detection of 3-mm signals generated in the laboratory.

[3] R. H. Dicke, Rev. Sci. Instr. **17**, 268 (1946).

[4] H. Zirin, Astrophys. J. **119**, 371 (1954).

Errata

Vibrational Anharmonicity and Lattice Thermal Properties, J. S. DUGDALE AND D. K. C. MACDONALD [Phys. Rev. **96**, 57 (1954)]. In the appendix to this paper, we attributed to Peierls and Domb the suggestion that a linear chain of atoms would be inherently unstable owing to thermal vibrations. Neither of these authors in fact implied this. They were concerned specifically with the existence of long range order as measured by x-ray diffraction in one-, two-,

and three-dimensional "crystals" and Peierls[1] established that such order persists only in three-dimensional crystals above the absolute zero of temperature. In the appendix to our paper we have demonstrated only that the relative atomic fluctuations remain finite, consistent with the thermodynamic stability of the chain (see also Frenkel[2]).

We are grateful to Professor Peierls and Professor Domb for discussions and correspondence on this topic.

[1] R. Peierls, Helv. Phys. Acta **81**, Suppl. 2 (1936).
[2] J. Frenkel, *Kinetic Theory of Liquids* (Clarendon Laboratory, Oxford, 1946), pp. 120–124.

Decay of Ti^{51} and Cr^{51}, M. E. BUNKER AND J. W. STARNER [Phys. Rev. **97**, 1272 (1955)]. In regard to paragraph III(d), one of the comparison factors used in the calculation of α_T was in error by a factor of 2.0. The corrected value is $\alpha_T=(1.62\pm0.16)\times10^{-3}$.[1] On the basis of this result, the K conversion coefficient is calculated to be $\alpha_K=1.47\times10^{-3}$. The two nearest theoretical conversion coefficients[2] are $\beta_K{}^1=1.1\times10^{-3}$, and $\alpha_K{}^2=3.8\times10^{-3}$. The 323-kev transition therefore appears to be $M1+E2$ with the $E2$ component having an intensity of $\sim13\%$.

[1] This value compares favorably with that recently reported by Z. O'Friel and A. H. Weber, Phys. Rev. **99**, 659(A) (1955); private communication. Their result is $\alpha_T=(1.48\pm0.2)\times10^{-3}$.
[2] Rose, Goertzel, Spinrad, Harr, and Strong, Phys. Rev. **83**, 79 (1951); Rose, Goertzel, and Swift (privately circulated tables).

Spin-Echoes with Four Pulses—An Extension to n Pulses, T. P. DAS AND D. K. ROY [Phys. Rev. **98**, 525 (1955)]. In this paper, the trigonometric parts of the amplitudes obtained by us for the echoes $P_{(12)}$, and the set $P_{(13)}$, $P_{(22)}$, $P_{((12)2)}$, $P_{(123)}$ (corresponding respectively to the primary echo produced by the first pair of pulses and the four secondary echoes produced by the first three pulses), are in disagreement with earlier values deduced by Hahn[1] and Das and Saha.[2] Our amplitudes are given in Table I together with those

TABLE I. Trigonometric part of amplitudes of several echo terms.

Term	Our amplitude	Hahn's amplitude
$P_{(12)}$	$\frac{1}{2}\sin^3\xi\cos^3(\xi/2)$	$\sin\xi\sin^3(\xi/2)$
$P_{(13)}$	$\frac{1}{4}\sin^3\xi\cos^3(\xi/2)$	$\frac{1}{4}\sin^3\xi$
$P_{(22)}$	$\frac{1}{4}\sin^3\xi$	$\sin\xi\sin^3(\xi/2)$
$P_{((12)2)}$	$-\frac{1}{4}\sin^3\xi\sin^3(\xi/2)$	$\sin\xi\sin^4(\xi/2)$
$P_{(123)}$	$\frac{1}{8}\sin^3\xi\sin^3(\xi/2)$	$\frac{1}{8}\sin^3\xi$

of the earlier workers for comparison ($\xi=\omega_1 t_w$, t_w being the width of the rf-pulse and $\omega_1=\gamma H_1$, where H_1 is the amplitude of the rf field during the applied pulse).

The cause of this disagreement is briefly as follows: In Hahn's analysis (see our paper and reference 1), to find the primary echo amplitude, only a single pair of pulses is applied and averaging[3] over $\Delta\omega$, η, and ϕ is done after the second pulse. Similarly for the secondary echoes, three pulses are applied and the averaging is done after the third pulse. In our paper, we were more interested in the echoes that follow after the fourth pulse. We therefore collected all the terms contributing to the V-component (the V-component in the rotating[1] system) of the nuclear magnetization, after the fourth pulse and then averaged over $\Delta\omega$, η, and ϕ. This gives The amplitudes of the echoes following the fourth pulse correctly but not those of the primary and secondary echoes produced by only the first two and the first three pulses respectively, as these echoes have their maxima before the fourth applied pulse (under the assumed condition $\tau_3>2\tau_2$), and only their tails remain after their fourth pulse (of course these tails cannot be seen because the echoes are limited to a width $1/T_2{}^*$ by the field inhomogeneity). If we are interested in the primary and secondary echo amplitudes we must therefore apply the averaging procedure to the terms contributing to V, after the second and third pulses respectively, when we get Hahn's result. We cite a similar disagreement between the two values obtained by Hahn[4] for the free-induction signal following the first pulse in the two cases when he analyzes the patterns following one and two pulses. He obtains the values $\sin\xi$ and $\sin\xi\cos^2(\xi/2)$ respectively, of which the former is the correct one.

[1] E. L. Hahn, Phys. Rev. **80**, 580 (1950).
[2] T. P. Das and A. K. Saha, Phys. Rev. **93**, 749 (1954).
[3] The symbols have the same significance as in reference 2.
[4] Refer to Eqs. (16) and (17) of reference 1.

Energy Levels of Li^6 from the Deuteron-Helium Differential Cross Sections, A. GALONSKY AND M. T. MCELLISTREM [Phys. Rev. **98**, 590 (1955)]. On page 598, second column, line 9, the definition of the Coulomb phase shift should read "α_l=Coulomb phase shift=$2[\arctan\eta+\arctan(\eta/2)+\cdots+\arctan(\eta/l)]\cdots$" instead of "$\alpha_l$=Coulomb phase shift=$2\arctan(\eta+\eta/2+\cdots+\eta/l)\cdots$".

Special-Relativistic Derivation of Generally Covariant Gravitation Theory, ROBERT H. KRAICHNAN [Phys. Rev. **98**, 1118 (1955)]. Equation (11) should read:

$$\mathfrak{D}^{\sigma\tau}(f)=|\eta|^{\frac{1}{2}}D^{\sigma\tau}(f)=\cdots \qquad (11)$$

instead of

$$\mathfrak{D}^{\sigma\tau}(f)=|\eta|^{\frac{1}{2}}\mathfrak{D}^{\sigma\tau}(f)=\cdots.$$

THE INDEX

TO THIS VOLUME HAS BEEN REMOVED

FROM THIS POSITION AND PLACED AT

THE BEGINNING OF THE FILM FOR THE

CONVENIENCE OF READERS.

and three-dimensional "crystals" and Peierls[1] established that such order persists only in three-dimensional crystals above the absolute zero of temperature. In the appendix to our paper we have demonstrated only that the relative atomic fluctuations remain finite, consistent with the thermodynamic stability of the chain (see also Frenkel[2]).

We are grateful to Professor Peierls and Professor Domb for discussions and correspondence on this topic.

[1] R. Peierls, Helv. Phys. Acta **81**, Suppl. 2 (1936).
[2] J. Frenkel, *Kinetic Theory of Liquids* (Clarendon Laboratory, Oxford, 1946), pp. 120–124.

Decay of Ti^{51} and Cr^{51}, M. E. BUNKER AND J. W. STARNER [Phys. Rev. **97**, 1272 (1955)]. In regard to paragraph III(d), one of the comparison factors used in the calculation of α_T was in error by a factor of 2.0. The corrected value is $\alpha_T=(1.62\pm0.16)\times10^{-3}$.[1] On the basis of this result, the K conversion coefficient is calculated to be $\alpha_K=1.47\times10^{-3}$. The two nearest theoretical conversion coefficients[2] are $\beta_K{}^1=1.1\times10^{-3}$, and $\alpha_K{}^2=3.8\times10^{-3}$. The 323-kev transition therefore appears to be $M1+E2$ with the $E2$ component having an intensity of $\sim13\%$.

[1] This value compares favorably with that recently reported by Z. O'Friel and A. H. Weber, Phys. Rev. **99**, 659(A) (1955); private communication. Their result is $\alpha_T=(1.48\pm0.2)\times10^{-3}$.
[2] Rose, Goertzel, Spinrad, Harr, and Strong, Phys. Rev. **83**, 79 (1951); Rose, Goertzel, and Swift (privately circulated tables).

Spin-Echoes with Four Pulses—An Extension to n Pulses, T. P. DAS AND D. K. ROY [Phys. Rev. **98**, 525 (1955)]. In this paper, the trigonometric parts of the amplitudes obtained by us for the echoes $P_{(13)}$, and the set $P_{(13)}$, $P_{(23)}$, $P_{((12)3)}$, $P_{(123)}$ (corresponding respectively to the primary echo produced by the first pair of pulses and the four secondary echoes produced by the first three pulses), are in disagreement with earlier values deduced by Hahn[1] and Das and Saha.[2] Our amplitudes are given in Table I together with those

TABLE I. Trigonometric part of amplitudes of several echo terms.

Term	Our amplitude	Hahn's amplitude
$P_{(13)}$	$\frac{1}{2}\sin^3\xi\cos^2(\xi/2)$	$\sin\xi\sin^2(\xi/2)$
$P_{(13)}$	$\frac{1}{2}\sin^3\xi\cos^2(\xi/2)$	$\frac{1}{4}\sin^3\xi$
$P_{(23)}$	$\frac{1}{4}\sin^3\xi$	$\sin\xi\sin^2(\xi/2)$
$P_{((12)3)}$	$-\frac{1}{2}\sin^3\xi\sin^2(\xi/2)$	$\sin\xi\sin^4(\xi/2)$
$P_{(123)}$	$\frac{1}{8}\sin^3\xi\sin^2(\xi/2)$	$\frac{1}{8}\sin^3\xi$

of the earlier workers for comparison ($\xi=\omega_1 t_w$, t_w being the width of the rf-pulse and $\omega_1=\gamma H_1$, where H_1 is the amplitude of the rf field during the applied pulse).

The cause of this disagreement is briefly as follows: In Hahn's analysis (see our paper and reference 1), to find the primary echo amplitude, only a single pair of pulses is applied and averaging[3] over $\Delta\omega$, η, and ϕ is done after the second pulse. Similarly for the secondary echoes, three pulses are applied and the averaging is done after the third pulse. In our paper, we were more interested in the echoes that follow after the fourth pulse. We therefore collected all the terms contributing to the V-component (the Y-component in the rotating[1] system) of the nuclear magnetization, after the fourth pulse and then averaged over $\Delta\omega$, η, and ϕ. This gives The amplitudes of the echoes following the fourth pulse correctly but not those of the primary and secondary echoes produced by only the first two and the first three pulses respectively, as these echoes have their maxima before the fourth applied pulse (under the assumed condition $\tau_3>2\tau_2$), and only their tails remain after their fourth pulse (of course these tails cannot be seen because the echoes are limited to a width $1/T_2{}^*$ by the field inhomogeneity). If we are interested in the primary and secondary echo amplitudes we must therefore apply the averaging procedure to the terms contributing to V, after the second and third pulses respectively, when we get Hahn's result. We cite a similar disagreement between the two values obtained by Hahn[4] for the free-induction signal following the first pulse in the two cases when he analyzes the patterns following one and two pulses. He obtains the values $\sin\xi$ and $\sin\xi\cos^2(\xi/2)$ respectively, of which the former is the correct one.

[1] E. L. Hahn, Phys. Rev. **80**, 580 (1950).
[2] T. P. Das and A. K. Saha, Phys. Rev. **93**, 749 (1954).
[3] The symbols have the same significance as in reference 2.
[4] Refer to Eqs. (16) and (17) of reference 1.

Energy Levels of Li^6 from the Deuteron-Helium Differential Cross Sections, A. GALONSKY AND M. T. MCELLISTREM [Phys. Rev. **98**, 590 (1955)]. On page 598, second column, line 9, the definition of the Coulomb phase shift should read "α_l=Coulomb phase shift$=2[\arctan\eta+\arctan(\eta/2)+\cdots+\arctan(\eta/l)]\cdots$" instead of "$\alpha_l$=Coulomb phase shift$=2\arctan(\eta+\eta/2+\cdots+\eta/l)\cdots$".

Special-Relativistic Derivation of Generally Covariant Gravitation Theory, ROBERT H. KRAICHNAN [Phys. Rev. **98**, 1118 (1955)]. Equation (11) should read:

$$\mathfrak{D}^{\sigma\tau}(f)=|\eta|^{\frac{1}{2}}D^{\sigma\tau}(f)=\cdots \qquad (11)$$

instead of

$$\mathfrak{D}^{\sigma\tau}(f)=|\eta|^{\frac{1}{2}}\mathfrak{D}^{\sigma\tau}(f)=\cdots.$$